Détection et conciliation d'erreurs intégrées dans un décodeur vidéo : utilisation des techniques d'analyse statistique

Brice Ekobo Akoa

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **NANO ELECTRONIQUE ET NANO TECHNOLOGIES**

Arrêté ministériel : 7 août 2006

Présentée par

« Brice EKOBO AKOA »

Thèse dirigée par **« Emmanuel SIMEU »** et
Co-encadrée par **« Fritz LEBOWSKY »**

Préparée au sein du **Laboratoire TIMA**
Dans **l'École Doctorale EEATS**

DÉTECTION ET CONCILIATION D'ERREURS INTÉGRÉES DANS UN DÉCODEUR VIDÉO: UTILISATION DES TECHNIQUES D'ANALYSE STATISTIQUE

Thèse soutenue publiquement le 31 Octobre 2014
devant le jury composé de :

M. Ioannis PARISSIS
Professeur à l'Université de Grenoble Alpes, Président,

M. Abbas DANDACHE
Professeur à l'Université de Loraine, Rapporteur,

M. Kosai RAOOF
Professeur à l'Université du Mans Rapporteur,

M. Fritz LEBOWSKY
Docteur et Ingénieur R&D chez STMicroelectronics, Co-encadrant,

M. Emmanuel SIMEU
Maître de Conférences HDR à l'Université de Grenoble Alpes, Directeur de Thèse.

DEDICACE

A ma chère et tendre épouse Vanessa,
Car nulle dédicace ne peut exprimer ce que je te dois,
pour ta patience, ton affection, ton soutien…
et à mon fils Nathan, tous deux trésors de bonté, de générosité et de tendresse,
en témoignage de mon profond amour et
ma grande reconnaissance « Que Dieu vous garde ».

A mes chers parents AKOA AKOA Théodore et MENGUE EKOBO Angèle, pour votre bienveillance, votre soutien et vos sacrifices.
A mes chers beaux-parents mama Thérèse et papa Abraham.

A mes chers frères et sœurs,
en témoignage de mes sincères reconnaissances pour les efforts qu'ils ont consenti dans l'accomplissement de mes études.
A mes beaux-frères et belles-sœurs pour leur compréhension et leur soutien.

Aux membres de l'église Chandelier (en particulier à Pascal et Fabienne BONNAZ), qui m'ont soutenu
dans leur prière et m'ont assisté moralement.
A tous mes amis, pour leur aide et leur soutien moral durant l'élaboration de ce travail.

Avant-propos et remerciements

Ce manuscrit présente les travaux réalisés tout au long de ma thèse de doctorat. Ma thèse vient marquer la fin de mes études universitaires, et donc de tout mon parcours scolaire. Elle a été réalisée au sein du laboratoire TIMA de l'Institut Nationale Polytechnique (INP) de Grenoble en partenariat avec l'équipe architecture de STMicroelectronics Grenoble.

Au terme de ce travail je tiens à remercier Monsieur Emmanuel SIMEU qui m'a donné l'opportunité de réaliser mes travaux de recherche au laboratoire TIMA et m'a guidé tout au long de ma thèse, me permettant ainsi de combiner les connaissances de tout un gabarit intellectuel acquis dans divers domaines scientifiques durant mes études universitaires, à la perspicacité et l'efficacité de ses directives pour former un tout. Je remercie également Monsieur Fritz LEBOWSKY qui m'a donné la possibilité d'effectuer ma thèse avec une équipe déterminée et solidaire. Je lui adresse aussi mes vifs remerciements pour son aide précieuse et sa clairvoyance au cours de ces trois années de recherche où il m'a assisté en tant qu'encadreur.

Je tiens aussi à exprimer l'honneur que me font les membres de mon jury de soutenance d'accepter de juger ce modeste travail.

Je voudrais enfin remercier mes collègues qui ont contribué à l'établissement de ces résultats de recherche. Il s'agit plus particulièrement de Rshdee, Mourad, Noureddine et Rafik; mais je pense également à Ke, Louay, Laurent, Asma et Martin.

Table des matières

Avant-propos et Remerciements **3**
Table des matières **5**
Table des figures **9**
Liste des tableaux **13**
Liste des abréviations **15**
Introduction générale **19**
1 Méthodes d'évaluation de la Qualité d'une Vidéo **23**
1.1 Introduction 23
1.2 Historique 24
1.3 Méthode subjective d'EQV : Notes données par des humains 25
1.3.1 Les mesures subjectives 25
1.3.2 Traitement des données de l'expérience 27
1.3.3 Résultats et Discussions 27
1.4 Méthode objective d'EQI : Evaluation « Aveugle » de la Qualité d'Image 27
1.4.1 Scènes statistiques des images corrompues 28
1.4.2 Evaluation de l'intégrité et de la véracité des images basée sur l'identification 29
1.4.3 Evaluation des Performances 30
1.4.4 En résumé 34
1.5 Méthode « Hybrides » : Métriques de Qualité basées sur le SVH 35
1.5.1 Description des métriques proposées 35
1.5.2 Test des Images et Expérience subjective 36
1.5.3 En résumé 37

1.6 Conclusion 37
2 Evaluation de la Qualité d'une Vidéo par classification 39
2.1 Introduction 39
2.2 Combinaison des classifieurs 40
2.2.1 Création et manipulation du vecteur de qualité 40
2.2.2 Mesure des Performances 42
2.3 OMQV basé sur la classification 43
2.3.1 Généralités sur l'OMQV à base de la classification 44
2.3.2 Conception de l'OMQV basée sur la classification 44
2.4 Implémentation de l'OMQV 52
2.4.1 Apprentissage du classifieur 52
2.4.2 Validation du classifieur 53
2.5 Conclusion 58
3 Evaluation de la Qualité d'une Vidéo par Réseaux de Neurones Artificiels 59
3.1 Introduction 59
3.2 Mesures de qualité d'image utilisant l'approche RNA 60
3.2.1 La méthode proposée 61
3.2.2 L'apprentissage et le test 62
3.2.3 Résultats expérimentaux 62
3.2.4 Résumé et synthèse 63
3.3 OMQV basé sur les RNA 63
3.3.1 Concepts de base 63
3.3.2 Conception de l'OMQV basé sur les RNAs 64
3.4 Implémentation de l'OMQV basé sur les RNAs 65
3.4.1 Modélisation et apprentissage du RNA 65
3.4.2 Validation et résultats de simulation 66
3.5 Conclusion 69
4 Evaluation de la Qualité d'une Vidéo par Régression Non Linéaire 71
4.1 Introduction 71
4.2 Etat de l'art: Utilisation de la régression dans le test et le contrôle 72

4.3 OMQV basé sur la Régression 73
4.3.1 Généralités de l'OMQV basée sur la régression non linéaire 74
4.3.2 Conception de l'OMQV basée sur la RNL 77
4.4 Modélisation de l'OMQV basé sur la RNL 79
4.4.1 Algorithme de régression 79
4.4.2 Résultats et Discussions 82
4.5 Conclusion 85
5 Méthodes de diagnostic et la correction d'artefacts visuels 87
5.1 Introduction 87
5.2 Généralités sur les artefacts 88
5.2.1 Principaux types d'artefacts 88
5.2.2 Détection d'Artefacts 92
5.3 Quelques Techniques de dissimulation d'Artefacts Visuelles 94
5.3.1 Elimination d'Artefacts avec un mode de sélection Perceptivement Optimisé . 95
5.3.2 Algorithme avancé de dissimulation d'Artefacts pour des intra-trames 96
5.3.3 Algorithme de dissimulation basé sur le décodage H.264/AVC non-normatif .. 97
5.3.4 Elimination des effets de bloc dans les Systèmes de Décodage Vidéo 98
5.3.5 Technique de convolution ou de déconvolution avec une gaussienne 99
5.4 Conclusion 103
6 Boucle de contrôle de la qualité du décodage 105
6.1 Introduction 105
6.2 Analyse et Paramétrage des erreurs du décodage MPEG 106
6.2.1 Analyse des Artefacts visuels dans le décodage MPEG 106
6.2.2 Paramétrage de l'artefact "bruit" 107
6.2.3 Paramétrage de l'artefact "flou" 109
6.2.4 Paramétrage de l'artefact "effets de bloc" 109
6.2.5 Paramétrage de l'artefact "suroscillations" 110
6.3 Priorisation et Elimination d'artefacts visuels 112
6.3.1 Priorisation des artefacts suivant leur degré d'impact sur la qualité 112

6.3.2 Algorithme de correction d'erreurs assisté par l'OMQV 114
6.4 Mesure et test de la qualité d'une image 123
6.4.1 Choix du seuil de qualité 123
6.4.2 Performances du système de validation des images 124
6.5 Discussions et Perspectives 125
6.5.1 Conception d'une boucle de correction des erreurs dans le décodage MPEG 125
6.5.2 Comparaison avec un outil commercialisé 126
6.6 Conclusion 127
Conclusion et Perspectives 129
Bibliographie 131
Annexes 137
Annexe 1 : Liste des publications 137
Conférences internationales 137
Conférence nationale 137
Annexe 2: Détection et classification d'artefacts de compression vidéo 139
A.1 Introduction 139
A.2 Analyse 140
A.3 Contribution essentielle 146
A.4 Conclusion 152

Table des Figures

Figure 0.1 Différentes sortes d'utilisation de l'OMQV. 20
Figure 1.1 Mesure alternative: Principe de l'EQV par mesure alternative 24
Figure 1.2 Echelle de la qualité continue à 5 points..... 26
Figure 1.3 Organigramme des étapes du processus de calcul des MOS 26
Figure 1.4 DIIVINE: identification suivi d'une évaluation de la qualité [46]. 30
Figure 1.5 Des versions déformées d'images de la base de données LIVE. 31
Figure 1.6 Organigramme des étapes de calcul du PSNR-HVS 35
Figure 1.7 Comparaison des corrélation du PSNR et du PSNR-SVH avec le MOS [20]..... 37
Figure 2.1 Couleurs associées aux 5 classes de qualité..... 42
Figure 2.2 Valeurs de SI et TI des vidéos de la base de données. 46
Figure 2.3 Dépendance de la qualité d'une vidéo sur le taux de perte de paquets. 48
Figure 2.4 Une frame "foreman" :originale, avec 0,4% de PLR et avec 5% PLR. 48
Figure 2.5 Classes de qualité : MOS vs. PSNR pour la phase validation. 49
Figure 2.6 Qualité de la vidéo vs. PSNR pour la séquence Vidéo "foreman"..... 50
Figure 2.7 Variation du niveau de flou en fonction du paramètre de floutage BT.. 51
Figure 2.8 Schéma bloc simplifié de l'estimation de la métrique flou. 51
Figure 2.9 Algorithme de classification d'une vidéo $v_{0.}$ 54
Figure 2.10 Vidéos d'apprentissage (rond) et validation (carré). 55
Figure 2.11 Résultats de classification pour la phase de test. 55
Figure 2.12 Erreurs quadratiques moyennes obtenu sur la classification en validation. 56
Figure 2.13 Taux de réussite du classifieur sur les 5 classes de qualité vidéo. 56
Figure 2.14 Classification obtenue pour les séquences vidéo de format 4CIF..... 57
Figure 3.1 Organigramme de la méthode proposée. 61
Figure 3.2 Modèle du RNA. 61
Figure 3.3 Architecture du réseau de neurones artificiels. 64
Figure 3.4 Fonctions d'activation usuelles. 65
Figure 3.5 Architecture du RNA implémenté pour l'OMQV. 66
Figure 3.6 Capture d'écran de l'évaluation d'une séquence vidéo par le RNA. 66

Figure 3.7 MOS vs. scores estimés par RNA à la fin de la phase d'apprentissage. 67
Figure 3.8 Corrélation entre MOS estimés et MOS pour la phase de test. 68
Figure 3.9 MOS vs. MOS estimés par le RNAs, pour une vidéo au format 4CIF. 68
Figure 4.1 Synoptique général de génération de test alternatif [57].. 73
Figure 4.2 Influence du PLR sur le MOS (en bleu) et les modèles de régression associés. 75
Figure 4.3 Dépendance du MOS sur le PSNR. 76
Figure 4.4 Dépendance du MOS (en ordonnées) en fonction de la métrique de flou. 77
Figure 4.5 Algorithme de prédiction du MOS estimé par régression non-linéaire. 78
Figure 4.6 MOS (en abscisse) vs Scores estimés par RNL (en ordonnées), 1ère itération. . 81
Figure 4.7 MOS (en abscisse) vs Scores estimés par RNL (en ordonnées), 2ème itération. 81
Figure 4.8 MOS (en abscisse) vs Scores estimés par RNL (en ordonnées), 3ème itération. 82
Figure 4.9 Vidéo "mobile" :de gche à dte: réf. ; avec PLR = 0.4% ; et PLR=5%..... 83
Figure 4.10 Scores estimés et MOS en l'absence des attributs SI et TI. 83
Figure 4.11 Classification et mesure par régression des vidéos de la base de données 84
Figure 5.1 Architecture d'un décodeur MPEG: Identification des sources d'artefacts. 89
Figure 5.2 Frame "Foreman":image d'origine, image flou et image restaurée..... 90
Figure 5.3 Image originale (a), absence de blocs b: Image jpeg compressée à 10%. 90
Figure 5.4 Trame "foreman" : A gauche : réf.; Au milieu :PLR 0.4%. A droite :PLR 5%... 91
Figure 5.5 Une image tirée touchée par des artefacts de bruit 92
Figure 5.6 Schéma bloc de l'algorithme proposé dans [5]..... 95
Figure 5.7 Les 8 sous-blocs adjacents au MB manquant 96
Figure 5.8 Sous-blocs correspondant dans la trame, et les MB candidats 96
Figure 5.9 PSNR moyen vs. PLR pour la séquence vidéo container 97
Figure 5.10 Trame décodée avec 20% par l'algorithme proposé en [60]. 98
Figure 5.11 Diagramme bloc de la dissimulation 3D de l'effet de bloc..... 98
Figure 5.12 Image CAMERAMAN: bruitée et restauré par filtrage bilateral 100
Figure 5.13 Image CAMERAMAN : originale et floutée par convolution avec PSF. 100
Figure 5.15 Images restaurées avec les filtresSOUSPSF; SURPSF ; et INITPSF.. 101
Figure 5.16 Aperçue des filtres de répartition PSF, SOUSPSF, SURPSF, INITPSF..... 102
Figure 5.17 Bords de l'image CAMERAMAN et image restaurée 102
Figure 6.1 Schéma bloc illustrant l'analyse d'erreurs dans le décodeur MPEG. 106
Figure 6.2 Détection et Analyse des erreurs dans le décodage MPEG. 107
Figure 6.3 Une trame mobile & calendar avec bruit blanc gaussien additif 108
Figure 6.4 PSNR en fonction de la variance du bruit blanc gaussien. 108
Figure 6.5 Une trame foreman avec flou gaussien..... 109
Figure 6.6 Correlation entre MOS et la métrique NPBM. Obtenu de [39]. 110
Figure 6.7 Artefacts générés dans une zone d'une trame de «mobile & calendar » 111
Figure 6.9 Variation du score de qualité visuelle en fonction du flou de l'image. 111
Figure 6.10 Variation du score de qualité visuelle en fonction du bruit de l'image..... 112
Figure 6.10 Graphe de priorisation des artefacts. 113
Figure 6.11 Correction d'erreurs assisté par l'OMQV dans un décodeur MPEG. 114
Figure 6.12 Schéma de la correction d'erreurs assistée par l'OMQV. 115
Figure 6.13 Schéma réel de la boucle de correction des images. 115
Figure 6.15 Correction d'erreurs dans le cas d'un bruit nul..... 116

Figure 6.16 Variations du niveau de flou de l'image en fonction de la taille *d* du PSF. 117
Figure 6.17 Variation du bruit de l'image en fonction de la taille *d* du masque PSF. 118
Figure 6.18 Variation du flou de l'image traitée avec l'écart-type σbdu bruit gaussien. 120
Figure 6.19 Variation de PSNR de l'image traitée avec l'écart-typeσb du bruit gaussien. . 120
Figure 6.20 Résltats de simulation : mesures des artefacts prises comme fonctions coûts.. 121
Figure 6.21 Résultats de dissimulation : scores estimés pris comme fonction coût. 122
Figure 6.23 Schéma de la boucle de contrôle de la qualité du décodage. 126
Figure A.0.1 Positionnement et fonctions du détecteur d'artefacts..... 140
Figure A.0.2 Compression des images par le décodeur JPEG. 141
Figure A.0.3 Réordonnancement en zigzag. 142
Figure A.0.4 SVM avec échantillons linéairement séparables..... 143
Figure A.0.5 SVM avec des échantillons non linéairement séparables. 143
Figure A.0.6 Schéma de classification à l'aide des SVM. 143
Figure A.0.7 Comparaison entre SVM avec parallélisme et SVM sans parallélisme..... 145
Figure A.0.8 Algorithme de calcul du vecteur représentatif. 146
Figure A.0.9 Valeurs du VS (sur l'axe y) pour 2 niveaux du bruit rose. 147
Figure A.0.10 Méthode de calcul de l'activité pour filtrages Passe-haut et Passe-bande..... 148
Figure A.0.11 Corrélation entre DMOS et métrique de netteté..... 149
Figure A.0.12 Schéma de calcul de la métrique objective du flou dans une image *I*..... 149
Figure A.0.13 Corrélation entre DMOS et la métrique du flou, base de données CSIQ. 150
Figure A.0.14 Corrélation entre métrique CPBD et DMOS. 151
Figure A.0.15 Corrélation CPBD vs DMOS, après normalisation..... 152

Liste des tableaux

Tableau 1.1 Valeurs du coefficient SROCC sur les échantillons de la BD LIVE. 33
Tableau 1.2 Le coefficient SROCC sur la base de données TID2008..... 34
Tableau 2.1 Comparaison des Performances des différentes méthodes [8]. 43
Tableau 3.1 Corrélation obtenue pour chaque type artefact. 62
Tableau 3.2 Corrélation obtenue pour tous les types d'artefact. 63
Tableau 3.3 Vecteur d'entrée (métriques) de la séquence ice avec un PLR de 10%. 67
Tableau 4.1 Métriques, MOS estimés et MOS pour les frames en Figure 4.9. 82
Tableau 6.1 Echelle à 5 niveaux de qualité visuelle des images. 123

Liste des abréviations

4CIF	4 times CIF (spatial resolution of 704x576 pixels)
ADSL	Asymmetric Digital Subscriber Line
ANOVA	ANalysis Of Variance
ASCII	American Standard Code for Information Interchange
AVC	Advanced Video Coding
CABAC	Context Adaptive Binary Arithmetic Coding
CHD	Connected Home Division
CIF	Common Intermediate Format (spatial resolution of 352x288 pixels)
CPBD	Cumulative Probability of Blur Detection
CSIQ	Content-Based Strategies of Image and Video Quality Assessment
CUT	Circuit Under Test
DCG	Digital Convergence Group
DCT	Discrete Cosine Transform
DIVIINE	Distortion Identification-based Image Verity and INtegrity Evaluation
DMOS	Differential MOS
DOG	Difference of Gaussian
DOOG	Difference of Offset Gaussian
EPFL	Ecole Polytechnique Fédérale de Lausanne
EQI	Evaluation de la Qualité d'une Image

EQM	Erreur Quadratique Moyenne
EQV	Evaluation de la Qualité d'une Vidéo
FEM	Fonction d'Estimation des MOS
FF	Fast Fading
FFT	Fast Fourrier Transform
fps	frame per second
FR	Full Reference
FRI	Filtre à Réponse Impulsionnelle
GBM	Gradient-based Boundary Matching
GSM	Gaussian Scale Mixture
HEVC	High Efficiency Video Coding
HVS	Human Visual System
IP	Internet Protocol
IQA	Image Quality Assessment
IQV	Indice de Qualité d'une Vidéo
ITU	International Telecommunication Union
JPEG	Joint Photographic Expert Group
JPEG2k / JPEG2000	The JPEG version introduced in 2000
kbps	kilo bit per second
k-NN	k Nearest Neighbors
LCC	Linear Correlation Coefficient
LIVE	Laboratory for Image & Video Engineering
MAD	most annoying distorsion
MAE	Mean Absolute Error
MB	Macrobloc
MCR	Moindres Carrés Récursifs
MLP	Multi Layer Perceptron

MOF	Métrique objective du Flou
MOS	Mean Opinion Scores
MPEG	Moving Picture Experts Group
MRM	Most Relevant Metric
MSE	Mean Squared Error
MS-SSIM	Multi-Scale Structural SIMilarity
MV	Motion Vector
NALU	Network Abstraction Layer Units
NPBM	No-Reference Perceptual Blockiness Metric
NR	No Reference
OMQV	Outil de Mesure de la Qualité d'une Vidéo
PLR	Packet Loss Rate
Polimi	Polytechnico de Milan
PSF	point-spread function
PSNR	Peak Signal to Noise Ratio
QA	Quality Assessment
QoS	Quality of Service
RAM	Random Access Memory
RF	Radio Frequency
RLE	Run Length Encoding
RMSE	Root Mean Square Error
RNA	Réseau de Neurones Artificiels
RNL	Régression Non Linéaire
RR	Reduced Reference
RTP	Real-time Transfer Protocol
SCQV	Système de Contrôle de la Qualité d'une Vidéo
SDTV	Standard Definition TV

SI	Spatial Index
SNR	Signal to Noise Ratio
SROCC	Spearman’s Rank Ordered Correlation Coefficient
SSIM	Structural Similarity Index Metric
Stdspace	Standard deviation in the space domain
SVH	System Visuel Humain
SVM	Support Vector Machines
TI	Temporal Index
TID	Tampere Image Database
TID2008	TID release in 2008
TPP	Taux de Perte de Paquets
VIF	Visual Information Fidelity
VQEG	Visual Quality Experts Group
VQI	Video Quality Index
VQMT	Video Quality Measuring Tool
VS	Vecteur (de) Support
WN	White Noise
YUV	Luma (Y) and Chroma (UV) components of image in the color space

Introduction générale

Le monde entier vit l'ère du tout numérique, où la vidéo et l'imagerie prennent une place primordiale dans la société, parce que devenus incontournable dans la communication. Dès lors, les nouveaux défis des chercheurs et ingénieurs du domaine de l'imagerie et de la vidéo numériques consistent à chercher une amélioration continue la qualité de service des équipements d'imageries et de vidéos numériques tels que les décodeurs.

Dans les décodeurs vidéo tels que MPEG2 ou MPEG4 les préoccupations principales d'optimisation sont traditionnellement orientées vers la conformité aux normes. Ceci impose des contraintes à respecter dans l'implémentation des algorithmes de correction d'erreurs embarqués dans ces décodeurs. Ces contraintes compliquent encore plus ces tâches visant à dissimuler les artefacts dans les images et de ce fait, les objectifs ne sont pas souvent atteints. En conséquence, différents artefacts visuels peuvent rester perceptibles après le processus de décodage. La plupart de ces artefacts démontrent des comportements non linéaires des systèmes en raison de la nature des méthodes modernes de codage/décodage vidéo. Il importe alors de trouver des méthodes adaptées pour la dissimulation de ce type d'artefacts dans le décodage numérique. C'est dans ce cadre qu'une idée - ayant fait l'objet du présent projet- a consisté à étudier et à explorer des techniques basées sur l'intelligence artificielle ou l'analyse et l'apprentissage statistique [59] pour la mesure, la correction et le contrôle du niveau de qualité des vidéo et des images dans des décodeurs numériques.

Une approche pour résoudre ce problème consiste à analyser les sources d'erreurs les plus importantes dans une correction proportionnelle au degré de détérioration, ce qui nécessite la détection et la priorisation. Vu les contraintes liées aux normes des standards, le défi est de trouver une méthode appropriée pour incorporer les méthodes de dissimulation d'artefacts visuelles dans la boucle du décodage vidéo numérique. Cela permettrait de détecter des artefacts apparaissant en aval ou au sein même du processus de décodage et d'en proposer une correction.

Comme nous voulons répondre au mieux aux attentes de l'œil humain, l'évaluation de la qualité d'une image par l'OMQV (Outil de Mesure de la Qualité d'une Vidéo), en anglais VQMT (Video Quality Monitoring Tool)) est comparée à l'évaluation subjective par des observateurs scores moyens (notés MOS (Mean Opinions Scores)) donnés par les humains au cours d'une expérience subjective. Pour le choix de la base de données subjective, nous avon étudié quelques unes disponibles au grand public [17-18], [50-51]. Nous avons trouvé utile de choisir une base de données avec un ensemble d'images couvrant un large éventail de niveaux de complexité de l'information spatiale (SI) et de l'information temporelle (TI). Cette base de

données intègre des vidéos où la perte de paquets a été introduite par simulation d'une transmission dans des réseaux propices aux erreurs de transmission. Cela nous permettra d'examiner les erreurs indépendantes de la source de capture des images, telles que des erreurs dans le réseau de transmission. Une base de données remplissant ces critères est la base de données EPFL [17-18]. Comme cette base de données fournit juste des MOS données sur les vidéos et pas sur les images, nous avons uniquement considéré les paramètres vidéo ne tenant pas compte les aspects temporels. Ainsi, pour une vidéo donnée contenant n trames (images), la valeur de chaque paramètre pi à l'entrée de l'OMQV est une moyenne statistique des n valeurs des paramètres pi sur les n images constituant la vidéo. Dans un autre aspect de cette manière, cette évaluation de la qualité d'une image ou d'une vidéo vise à être en corrélation avec plusieurs paramètres caractéristiques différents et l'OMQV est conçu pour extraire les paramètres à partir d'une image ou d'une vidéo donnée.

Dans cette étude, la correction sera réalisée dans le cadre spatial de la vidéo. Une séquence vidéo étant constituée de plusieurs images qui se succèdent dans le temps. Mais afin de simplifier le travail, les aspects associés à la dynamique temporelle seront omis dans cette première phase du projet. Les techniques de correction d'erreurs présentées dans la deuxième et dernière partie de ce manuscrit se limiteront à la correction d'erreurs dans une image.

La principale contribution de ce travail est la possibilité de contrôler le niveau de qualité d'un système multimédia de façon à assurer une haute qualité des vidéos décodées vis-à-vis de la perception visuelle humaine. Cela a conduit à l'idée d'évaluer les méthodes utilisées dans l'analyse statistique avancée. La Figure 0.1 représente deux applications différentes de l'utilisation de l'OMQV dans les réseaux multimédias: application A - surveillance de réseau [52] et l'application B- correction des erreurs de décodage ou contrôle de la qualité d'un processus décodage. Nous envisageons de concevoir et d'implémenter un OMQV qui soit applicable dans les deux cas.

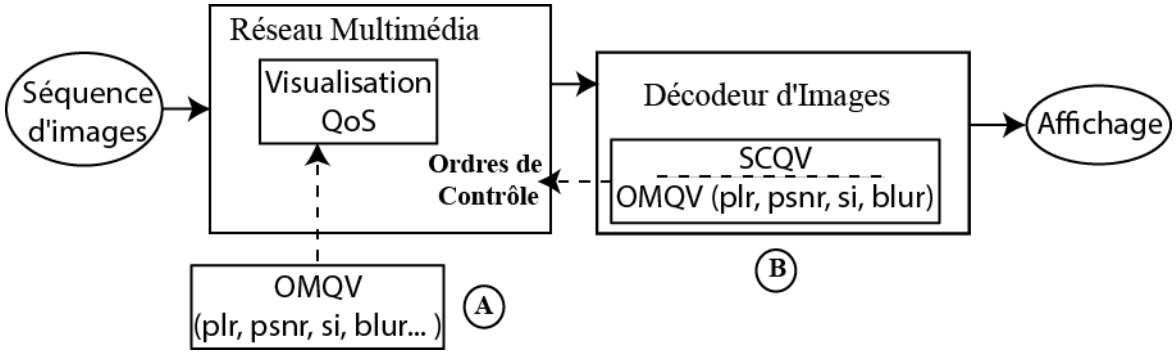


Figure 0.1 Différentes sortes d'utilisation de l'OMQV.

La problématique principale abordée ici est donc le contrôle de la qualité du décodage numérique de la famille MPEG. La notion de qualité visée est celle du jugement humain sur la qualité d'une image. Le défi est alors sur deux points principaux: Elaborer une bonne méthode pour superviser un processus de décodage numérique pour les contraintes d'architecture et assurer la conformité des résultats obtenus vis-à-vis du système visuel humain.

Le manuscrit comprend six chapitres et est structuré de la façon suivante : Les quatre premiers chapitres présentent une étude sur les solutions d'évaluation de la qualité des vidéos (EQV) basées sur de l'analyse statistiques avancées ou sur l'intelligence artificielle. Trois

outils de mesure de la qualité d'une vidéo (OMQV) sont proposés. Les deux derniers chapitres présentent notre proposition pour le contrôle de la qualité d'une vidéo au sein d'un décodeur MPEG.

Le premier chapitre présente les principales méthodes d'évaluation de la qualité des images trouvées dans l'état de l'art et introduit notre proposition.

Trois modèles de VQMT sont conçus et modélisés. Ils sont basés sur la classification, les réseaux de neurones artificiels et la régression non linéaire, et sont développés dans le deuxième, troisième et quatrième chapitre respectivement.

Le cinquième chapitre présente quelques techniques de dissimulation d'artefacts présents dans l'état de l'art.

Le sixième et dernier chapitre utilise les résultats des quatre premiers chapitres pour mettre au point un algorithme de correction d'erreurs dans les images.

Enfin nous finiront par une conclusion et des perspectives afin de résumer le manuscrit, présenter les résultats obtenus et montrer les différentes pistes évoquées en lien avec la thèse pour d'éventuels futurs projets.

1 Méthodes d’évaluation de la Qualité d’une Vidéo

1.1 Introduction

Avec le développement des récentes technologies des réseaux multimédias, le niveau d’exigence en qualité de service des systèmes multimédia s’accroit rapidement. L’optimisation des décodeurs est nécessaire pour garantir une meilleure qualité multimédia, conformément au système visuel humain. Le codage de données dans certains cas peut introduire des éléments tels que le bruit de codage/décodage et des erreurs lors du traitement de l'information. Ces éléments font apparaître des artefacts visuels qui détériorent la qualité des images. Il est donc nécessaire de dissimuler les artefacts visuels à la sortie des images. Cela implique aussi d’identifier les sources possibles d’artefacts visuels à la sortie des décodeurs numériques. Pour détecter ces sources d'erreur il est astucieux de d'abord mesurer la qualité du décodage en question. Cette nécessité a conduit les chercheurs à développer de grands intérêts pour l'application de techniques avancées d’intelligence artificielle pour l'évaluation de la qualité des vidéos (EQV). De la même façon, il est important qu’une EQV fournisse des résultats conformes au jugement humain pour aider efficacement à améliorer la qualité des vidéos ou des images.

On classe les méthodes d’EQV suivant deux différents critères. Le critère suivant le degré d’intervention de l’homme dans la mesure de la qualité ou le critère lié à l’intensité d’informations que requière la mesure sur l’image, ou la vidéo de référence. On compte ainsi une grande diversité de méthodes d’EQV ou d’EQI (évaluation de la qualité d’une image) dans l’état de l’art. En considérant toute cette variété de méthodes d’EQV, toute la problématique est de trouver la bonne méthode qui soit la moins dépendante de la vidéo ou de l’image de référence et qui procure des résultats les plus proches possible du jugement humain.

Le reste de ce chapitre est organisé comme suit : le paragraphe 2 présente un bref historique de la mesure de la qualité des images ou des vidéos; le paragraphe 3 présente une méthode subjective pour d’EQV, le paragraphe 4 décrit quelques méthodes objectives d’EQV tandis que le paragraphe 5 conclut par un résumé et une brève synthèse des différentes techniques.

1.2 Historique

Les premières méthodes d'EQV sont essentiellement « subjectives » [47] et se basent sur des notes données par des humains. Des séquences vidéo sont regardées par un groupe d'observateurs qui les notent ensuite suivant une échelle donnée. Le score (note finale) d'une séquence vidéo appelée MOS est ensuite calculé comme la moyenne des notes données par les observateurs. Le MOS représente donc un indice de qualité de la vidéo (IQV) en anglais VQI (Video Quality Index) (Figure 1.1).

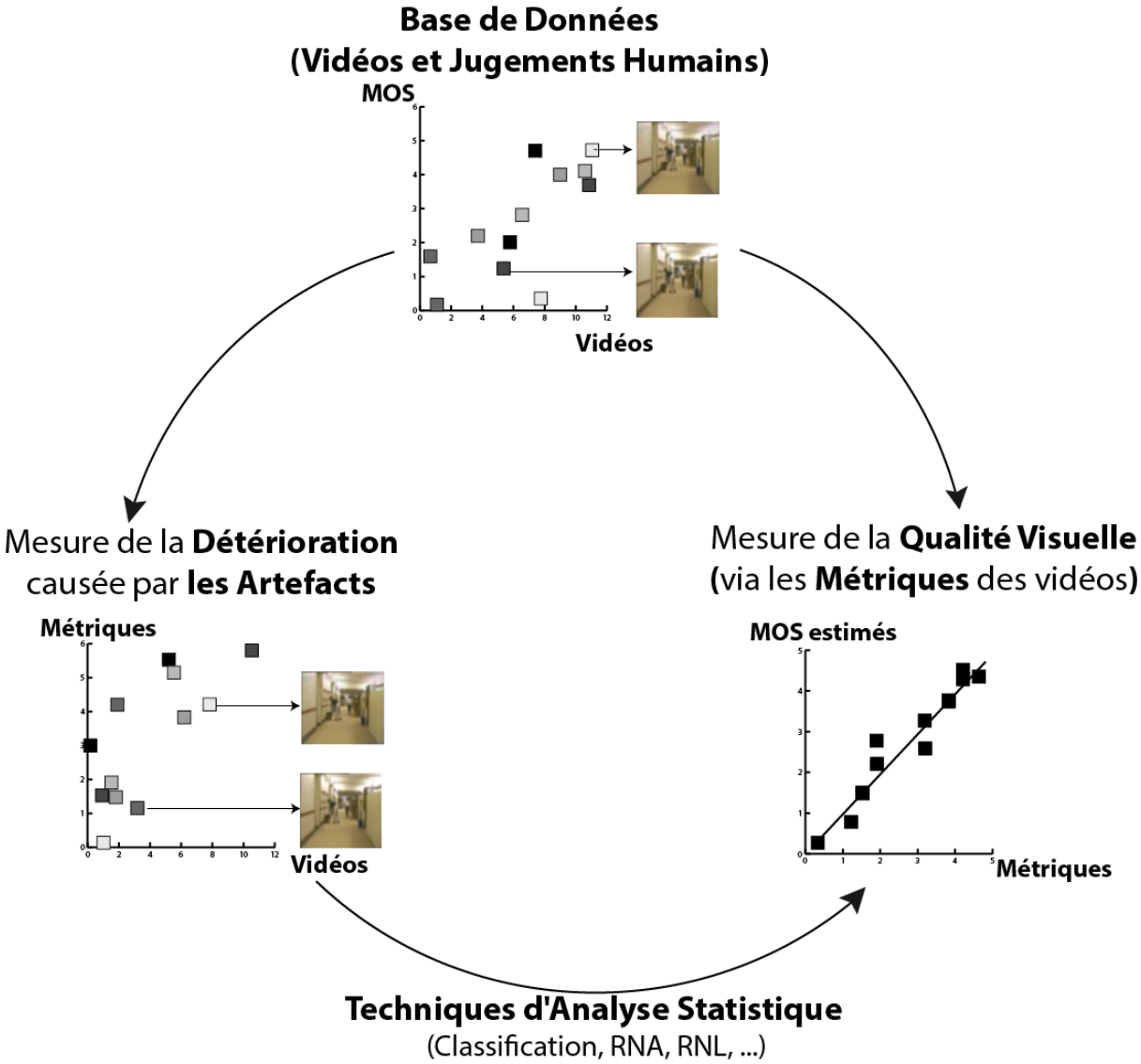


Figure 1.1 Mesure alternative: Principe de l'EQV par mesure alternative

Cependant, cette méthode ne convient pas pour des scénarios d'application en temps réel nécessitant des tâches répétitives d'évaluation continue de la qualité d'une séquence vidéo. Certains chercheurs ont donc proposé des solutions pour l'EQV fidèlement au système visuel humain (corrélés aux MOS) par des méthodes d'EQV dites « objectives ». Parmi ces méthodes d'EQVs objectives, on peut distinguer les techniques dites EQVs à référence complète qui requièrent la version originale de la séquence vidéo pour l'estimation du niveau de qualité. Une deuxième classe de méthodes regroupe les techniques pour lesquelles seule une partie de l'information sur la séquence vidéo originale est requise, dites méthodes d'EQVs à référence partielle [37]. La dernière catégorie d'EQVs dite sans référence [3] regroupe les méthodes d'EQV qui nécessitent uniquement des informations de distorsion sur la vidéo à évaluer. Des travaux de recherche décrits dans [19] ont utilisé des tests subjectifs pour prouver la corrélation entre les images compressées et l'activité spatiale. Dans [19], des expériences subjectives ainsi que l'évaluation par le calcul des MOS sont utilisées pour prouver l'efficacité de 2 nouvelles métriques de qualité (Figure 1.1) qui tiennent compte de la perception visuelle humaine.

1.3 Méthode subjective d'EQV : Notes données par des humains

Les références [17] et [18] présentent un travail de recherche qui vise à fournir une base de données accessible au public contenant des MOS recueillies au cours d'expériences subjectives effectuées dans les locaux de 2 établissements d'enseignement : Polytecnico de Milan (Polimi) – en Italie et l'Ecole Polytechnique Fédérale de Lausanne (EPFL) – en Suisse. D'autres travaux de recherche [20-23] ont déjà fait état de tels éléments pouvant servir de base de référence pour des chercheurs dans le domaine de la qualité des vidéos, mais aucun de ces travaux n'a été publiquement disponible. Les seules bases de données disponibles au grand public disposant de résultats d'Expériences sur l'évaluation subjective de la qualité d'une vidéo connue jusque là étaient la base de données LIVE et TID2008 [35] pour les images en définition standard et EPFL pour les images en haute résolution. Le papier [17] fait une extension au format vidéo 4CIF par rapport au travail établi dans [16].

1.3.1 Les mesures subjectives

Une expérience [17] a consisté à soumettre 78 séquences vidéo CIF (Common Intermediate Format) sous l'appréciation (l'évaluation) de 40 individus. La disponibilité de MOS a permis la validation et la mise en place d'une base de données de référence. Cette base permet désormais d'établir une comparaison objective des systèmes d'évaluation de la qualité d'une vidéo, de manière à soutenir des résultats de recherche reproductibles.

Dans cette expérience, six séquences vidéo d'essai (six scènes différentes) ont été considérées au format CIF, choisi de sorte que les différents niveaux de l'information spatiale SI et de l'information temporelle TI puissent être représentés. Deux autres séquences distinctes des 6 premières ont été utilisées pour l'entraînement des individus sujets de l'expérience. Le codeur H.264/AVC et la version 14.2 du logiciel H.264/AVC de référence ont été utilisés pour générer les flux compressés des vidéos. Toutes les séquences ont été codées à l'aide du logiciel H.264/AVC en profil HAUT, pour obtenir des B-frames et par Codage Arithmétique Binaire en contexte adaptatif (CABAC) pour un codage plus efficace. Les séquences vidéo ont été générées suivant six taux de perte de paquets (PLR) différents [0,1%, 0,4%, 1%, 3%, 5%, 10%] pour simuler des erreurs en rafale. Chaque train binaire est décodé par le décodeur H.264/AVC intégrant la méthode de correction d'erreurs basée sur la

compensation du déplacement.

L'expérience subjective considère six séquences vidéo de résolution 352x288 pixels nommées Foreman, Hall, Mobile, Mother, News et Paris. La cadence des séquences d'image est de 30 fps, les vidéos originales sont disponibles en format progressif brut. Chaque trame est divisée en un nombre fixe de tranches. Chaque tranche est constituée d'une rangée complète de Macroblocs. Le contrôle du débit a été désactivé parce qu'il introduit des fluctuations de la qualité visuelle au cours du temps pour certaines séquences vidéos. Il a été remplacé par un paramètre de quantification fixe soigneusement sélectionné pour chaque séquence de manière à assurer une grande qualité visuelle en l'absence des pertes de paquets. Le paramètre de quantification a été réglé pour chaque séquence afin de ne pas dépasser un débit de 600 kbps, pouvant être considéré comme limite supérieure pour la transmission du format vidéo CIF sur les réseaux IP. Chaque séquence vidéo a été visuellement testée afin de voir si le paramètre de quantification choisi minimise les artefacts d'effet de bloc pouvant être induits par un codage avec pertes de paquets.

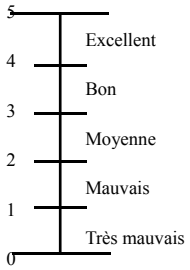


Figure 1.2 Echelle de la qualité continue à 5 points

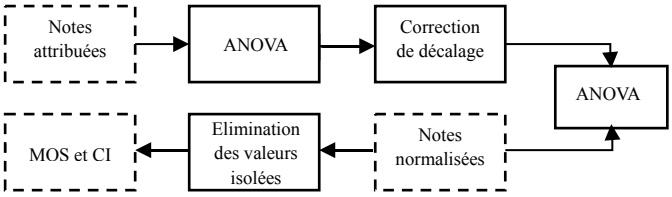


Figure 1.3 Organigramme des étapes du processus de calcul des MOS

Un contrôle précis et une description de l'environnement de test a été nécessaire pour assurer la reproductibilité de l'activité de test et de comparer les résultats entre les différents laboratoires et des séances d'essais. Le système d'éclairage ambiant des deux laboratoires était composé de néon lampes avec une température de couleur de 6500 K. Chaque séquence de test ne concerne qu'un seul individu par projection de vidéo. Ce qui a aussi permis d'évaluer le matériel de test, l'individu étant assis directement en ligne avec le centre de l'écran vidéo à une distance spécifiée. Chaque séquence est projetée pendant 10 secondes. A la fin de chaque

présentation, suit un temps de 3 à 5 secondes d'Evaluation. Quand un sujet évalue la qualité du stimulus en utilisant l'échelle continue à 5 points du groupe ITU-T dans l'intervalle [0 ; 5] illustré à la Figure 1.2. Les individus notent une séquence en mettant une croix sur l'échelle, qui va correspondre à une note calculée numériquement sur ordinateur. Les niveaux de l'échelle de qualité sont interprétés comme suit :

• Excellent: « le contenu de la séquence peut paraître un peu flou, mais ne présente aucun autre artefact visible»;
• Bon: « au moins un artefact est détecté lors de la lecture de la séquence vidéo»;
• Moyen: «plusieurs artefacts visibles sont détectés, répartis sur toute la séquence vidéo ».
• Mauvais: « de nombreux objets et des artefacts visibles solides (c.-à-d. les artefacts qui détruisent la structure de la scène ou qui créent de nouveaux modèles) sont détectés »;
• Très mauvais: "artefacts (artefacts puissants à savoir qui détruisent la scène structurer ou de créer de nouveaux modèles) sont détectés dans la majeure partie de la séquence ".

1.3.2 Traitement des données de l'expérience

La Figure 1.3 présente l'organigramme du processus de traitement des données subjectives. Une analyse de variance (ANOVA) a été effectuée dans le but de comprendre si une normalisation des scores serait nécessaire. Les résultats de l'ANOVA ont montré que la différence dans les moyennes entre les notes subjectives d'un sujet à un autre était grande. Ce qui signifie qu'il y avait des différences significatives entre les façons dont les sujets ont utilisé l'échelle de notation. Ainsi, une correction de sujet à un autre a été appliquée, en normalisant toutes les notes afin de compenser la moyenne des scores [19]. Le MOS a été calculé pour chaque condition de test en rejetant les notes isolées. L'intervalle de confiance de 95% et la distribution du T de Student ont également été calculés. Ces données statistiques ont tous été utilisées pour calculer les MOS à partir des scores moyens subjectifs obtenus.

1.3.3 Résultats et Discussions

Cette expérience a permis de produire une base de données publique sur les résultats d'une expérience subjective concernant la notation de 78 séquences vidéo au format CIF. Les taux subjectifs couvrent uniformément toute la gamme des niveaux de qualité. De plus, les intervalles de confiance sont de taille raisonnable. Ceci prouve que les efforts fournis par chaque observateur étaient relativement appropriés et que les observateurs étaient globalement assez cohérents dans leurs évaluations. Une extension à ces travaux pour le format vidéo 4CIF a été apportée par l'article décrit dans [18].

Dans les prochaines sections, des méthodes objectives d'EQV tirées de récents travaux de recherche sont décrites, faisant toute référence à une comparaison avec des MOS pour la validation des performances des métriques et techniques établies.

1.4 Méthode objective d'EQI : Evaluation « Aveugle » de la Qualité d'Image

Le terme "aveugle" pour une méthode d'EQI [45] désigne une méthode ne nécessitant aucune information sur l'image de référence ni une quelconque image du voisinage spatial.

L'expérience décrite en [46] part du fait que les images naturelles non déformées possèdent certaines propriétés statistiques qui tiennent à différents aspects de leurs contenus. L'approche d'EQI sans références proposée ici se fonde sur l'hypothèse que la présence de distorsions dans les images naturelles modifie la nature des propriétés statistiques de ces images. Ces distorsions enlève l'aspect naturelle de ces images (et par conséquent leurs statistiques). Dans un premier temps, les scènes statistiques extraites d'une image naturelle déformée sont utilisées explicitement pour classer l'image déformée dans l'un des n types de distorsion (identification de distorsion - étape 1). Puis, ces scènes statistiques vont servir à évaluer la qualité de cette distorsion spécifique (EQI de la distorsion spécifique - étape 2) de l'image. Une combinaison des deux étapes aboutit à un résultat de la qualité de l'image. L'approche proposée dénommée Evaluation de la Véracité et de l'Intégrité de l'image basée sur l'Identification des Distorsions, est notée DIVIINE (Distortion Identification-based Image Verity and INtegrity Evaluation).

1.4.1 **Scènes statistiques des images corrompues**

Le traitement de DIIVINE (Figure 1.4) se déroule de la façon suivante : dans un premier temps l'image est d'abord décomposée suivant une orientation basée sur l'échelle spatiale (faiblement, une transformée en ondelettes) pour former des réponses orientés en passe-bande. Les coefficients obtenus en bande de base sont ensuite utilisés pour extraire des paramètres statistiques. Ces paramètres forment alors un vecteur qui va constituer la description statistique du niveau de distorsion dans l'image. Dès lors, le but de la manœuvre est d'utiliser ce vecteur de paramètres statistiques pour procéder à deux traitement dans une séquence : (1) Identifier la probabilité que l'image soit affectée par une des catégories de distorsions préalablement considérées ; puis (2) affecter un score de qualité au vecteur de paramètres suivant chaque catégorie de distorsion. Construire ensuite un modèle de régression pour chaque catégorie de distorsion de manière à faire correspondre le vecteur paramètre statistique à un niveau de qualité conditionné par le fait que l'image soit affectée par cette distorsion. L'estimation de l'identification de distorsion probabiliste est ensuite combinée avec le niveau de qualité sans distorsion spécifique pour produire une valeur finale de la qualité de l'image.

Dans la première étape de la technique DIIVINE, un ensemble de coefficients d'ondelettes voisins est modélisé en utilisant un modèle de mélangeur à l'échelle gaussienne appelé GSM (Gaussian Scale Mixture).

Un vecteur aléatoire Y de dimension N est dit GSM si

$$Y \equiv z.U \tag{1.1}$$

où l'opérateur ≡ désigne l'égalité de probabilité de la distribution, U est un vecteur aléatoire de moyenne gaussienne nulle de covariance CU et Z une variable aléatoire scalaire appelé multiplicateur du mélangeur.

Afin d'illustrer la façon dont chacune de ces fonctions se comporte dans une image naturelle déformée, des images de référence non naturelles déformées ont été utilisées, et des images de synthèse distordues et crées à partir des images de références suivants les catégories de distorsions suivantes : la compression JPEG, la compression JPEG2000 (JP2k), le bruit blanc additif WN (White Noise), le Flou gaussien blur et un canal de Rayleigh étiqueté évanouissement rapide FF(Fast Fading).

Étant donné une image dont la qualité doit être évaluée, la première étape consiste à effectuer une décomposition en ondelettes en utilisant une pyramide « orientable » au-delà de 2 échelles et 6 orientations. Il a été constaté qu'un degré accru de sélectivité d'orientation est bénéfique à de fins de l'Evaluation de la Qualité (EQ) - plus que la sélectivité sur plus de 2 échelles. La décomposition de la pyramide orientable a été appliquée pour extraire des statistiques à partir d'images déformées [58].

.

1.4.2 **Evaluation de l'intégrité et de la véracité des images basée sur l'identification des distorsions**

L'approche DIIVINE pour l'EQI sans référence comprend 2 étapes. Elle consiste à utiliser les caractéristiques extraites de l'image corrompue comme décrit ci-dessus aussi bien pour l'identification du type de distorsion (ou d'artefact) que pour l'évaluation de la qualité du défaut (de la distorsion) spécifique. Ces deux étapes nécessitent un processus d'étalonnage qui relie chaque paramètre statistique calculé à la catégorie de distorsion et au MOS qui lui est associé. Cet étalonnage est réalisé par apprentissage avec un ensemble d'images (y compris les images corrompues suivant toutes les catégories de distorsion) et leurs MOS associés. Étant donné cet ensemble d'apprentissage et la calibration de ces deux étapes – identification de la distorsion et évaluation de la qualité de la distorsion spécifique - DIIVINE est capable d'évaluer la qualité de toute image déformée sans nécessité de l'image de référence. La phase d'étalonnage ne nécessite pas non plus d'image de référence.

L'expérience a considéré une suite d'images dédiées à l'apprentissage couvrant toutes les catégories de distorsion pour lesquelles l'algorithme est calibré. L'apprentissage de la classification a été effectué en considérant comme entrées les classes des images originales et les vecteurs paramètres statistiques. Durant l'apprentissage, la classification mémorise la correspondance entre les paramètres spatiaux de l'image et les labels des classes. Une fois l'apprentissage achevé, le « classifieur » entraîné produit une estimation de la classe de distorsion d'une image donnée.

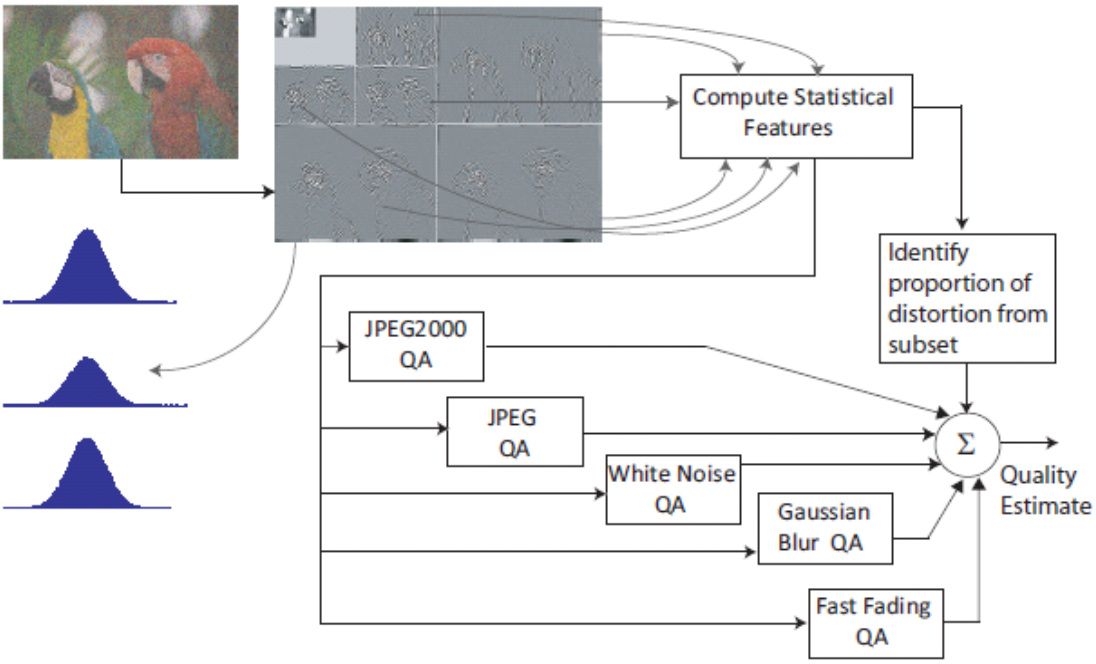


Figure 1.4 L'algorithme DIIVINE se compose de deux phases: distorsion probabiliste identification suivi d'une évaluation de la qualité de distorsion spécifique comme illustré ici [46].

Une suite d'images dédiées à l'apprentissage et les scores de qualité correspondants suivant chaque catégorie de distorsion a été mise en place. A partir d'elle, n modules de distorsions ont été entraînés pour faire correspondre un vecteur de paramètres statistique donné au score de qualité associé. Comme chaque module est entraîné pour une catégorie de distorsion spécifique, ces modèles de régression une fois formés constituent des évaluateurs de qualité d'image spécifique à une catégorie de distorsion.

Dans cette approche, le classifieur ne produit pas une classification stricte. Les estimations de probabilité sont plutôt extraites du classifieur, qui indique l'efficacité du classifieur mis en place pour classer chaque entrée dans chacun des n classes. Ainsi, étant donné une image / vecteur de paramètre d'entrée, le classifieur produit un vecteur de n dimensions, qui représente les probabilités que l'entrée appartienne à chacune des n classes.

1.4.3 **Evaluation des Performances**

La base de données utilisée

L'indice DIIVINE a été testé sur la base de données publique LIVE. Elle se compose de 29 images de référence et de 779 images déformées qui couvrent différentes catégories (voir Figure 1.5) de distorsion. Les notes différentielles DMOS associées aux scores moyens des observateurs MOS (Mean Opinion Scores) représentent la qualité perceptuelle de l'image.

Figure 1.5 Des versions déformées d'images de la base de données LIVE (a) - (e) correspondant aux distorsions suivantes - (a) la compression JP2k, (b) la compression JPEG, (c) bruit blanc, (d) Flou gaussien et (e) rapidement distorsion d'évanouissement.

L'algorithme DIIVINE nécessitait une phase d'apprentissage afin de calibrer la relation entre les caractéristiques statistiques et des extraits la catégorie de distorsion, ainsi que DMOS. Il convenait alors de diviser la base de données LIVE en 2 sous-ensembles disjoints -

un premier sous-ensemble pour l'entraînement et un second sous-ensemble de test. L'ensemble d'entraînement se compose de 80% des images de référence et de leurs versions déformées associées. L'ensemble de test comprend les 20% restants des images de référence et de leurs versions déformées associées. Les modules de classification et de régression sont formés sur le sous-ensemble de l'apprentissage et les résultats sont ensuite testés sur le sous-ensemble de test. Afin d'assurer que l'approche proposée est robuste à travers le contenu et n'est pas influencé par la scission de apprentissage-test spécifique utilisé, cet apprentissage aléatoire a été répété 1000 fois sur l'ensemble de la base de données LIVE et évaluer la performance de chacun des ces tests fixe.

Les indices utilisés pour mesurer les performances de l'algorithme sont : le Coefficient de corrélation de rangs de Spearman SROCC (Spearman's Rank Ordered Correlation Coefficient) pour vérifier la monotonie ; Le PCC (Pearson Corrélation Coefficient) ou coefficient de corrélation de Pearson pour la pertinence des résultats par comparaison avec d'autres (avec les MOS par exemple) ; Le coefficient de corrélation linéaire LCC (Local Correlation Coefficient) et la racine de l'erreur quadratique moyenne ou RMSE (Root Mean Squared Error) entre le score prédit et les DMOS. La précision d'une prédiction est quantifiée par le PCC et la RMSE [12]. Le SROCC mesure la monotonie de prédiction d'une métrique. Avant de calculer ces coefficients de corrélation, il est habituel d'appliquer une transformation non linéaire sur les scores prédits de manière à amener les prédictions sur la même échelle que
celle des scores subjectifs. Cette étape permet d'obtenir une relation linéaire entre les prédictions et les scores d'opinion comme dans [24]. Une valeur proche de 1 pour le SROCC et le LCC et une valeur proche de 0 pour RMSE indiquent une corrélation conforme à la perception humaine. Les valeurs du coefficient SROCC à travers les 1000 échantillons d'entraînement et de tests choisis sont représentées dans le tableau 1.1, pour chaque catégorie de distorsion.

Tests à caractères statistiques

L'analyse ici est basée sur les valeurs SROCC dans toutes distorsions. Il faut rappeler que les corrélations ont été calculées pour chacun des algorithmes plus de 1000 jeux de tests. Ainsi, en dehors des scores de médiane compilés avant, la moyenne des coefficients SROCC est aussi connue ainsi que l'erreur-type associée avec les 1000 des valeurs de corrélation. Cette valeur moyenne des corrélations est ensuite tracée à travers l'ensemble des données avec des barres d'erreur d'une norme déviation large.

En outre, DIIVINE est statistiquement plus performant que la métrique à référence complète PSNR (1.2). En effet, DIIVINE est non seulement capable d'évaluer la qualité dans de nombreuses catégories de distorsion, mais effectue également de bien meilleurs résultats statistiques comparés au PSNR. DIIVINE prédit la qualité visuelle d'une image déformée en corrélation avec le jugement humain à un niveau qui est statistiquement indiscernable de l'indice de similarité structurelle SSIM (Structural Similarity Index Metric). Le SSIM a cependant lui besoin à la fois de l'image de référence et de l'image déformée afin d'évaluer la qualité! Cela donne à penser que l'on peut remplacer en toute sécurité le FR SSIM avec le DIIVINE NR sans aucune perte de performance, à condition que les distorsions rencontrées soient bien représentées par l'ensemble de données utilisé pour former DIIVINE.

Tableau 1.1 Valeurs du coefficient SROCC sur les 1000 échantillons d'Apprentissage et test de la base de données LIVE. Les algorithmes en Italique sont sans référence, et les autres sont des algorithmes à référence complète.

	JP2K	JPEG	WN	Gblur	FF	All
PSNR	0.868	0.885	0.943	0.761	0.875	0.866
SSIM (SS)	0.938	0.947	0.964	0.907	0.940	0.913
BIQI-PURE	0.736	0.591	0.958	0.778	0.700	0.726
BIQI-4D	0.802	0.874	0.958	0.821	0.730	0.824
Anisotropic	0.173	0.086	0.686	0.595	0.541	0.323
BLIINDS	0.805	0.552	0.890	0.834	0.678	0.663
DIIVINE	0.913	0.910	0.984	0.921	0.863	0.916

Indépendance de la base de données

Les algorithmes EQI sont généralement entraînés et testés sur divers sous-ensembles d'un seul ensemble de données (tel que décrit dans ce projet). Il est alors naturel de se demander si l'ensemble formé de paramètres sont spécifiques à des bases de données. Afin de démontrer que le processus de formation est simplement un étalonnage et qu'une fois l'apprentissage effectué DIIVINE est capable d'évaluer la qualité de toute image déformée, sa performance a été évaluée sur une autre base de données – le TID2008. La base de données TID est composée de 25 images de référence et 1700 images déformées de plus de 17 catégories de distorsion. Parmi ces 25 images de référence, seuls 24 sont des images naturelles. L'algorithme est testé uniquement sur ces 24 images. En outre, sur les 17 catégories de distorsion, DIIVINE est uniquement testé sur les catégories pour lesquelles il a été entraîné - JPEG, JPEG2000 compression (JP2k), Bruit blanc additif (WN) et Gaussian Blur (flou). Afin d'évaluer DIIVINE sur la base TID, les paramètres de DIIVINE sont entraînés en utilisant la base de données entière LIVE. Le modèle entraîné est ensuite testé pour sa performance sur la base TID. La présente les valeurs du coefficient SROCC obtenues pour ces tests pour chaque type de distorsion et établit également la liste des performances des métriques PSNR et SSIM pour comparaison.

Analyse de la complexité des calculs

Bien que DIIVINE n'a pas été développé sous la contrainte d'analyse en temps réel d'images, compte tenu de ses performances aussi satisfaisants que ceux des principaux

algorithmes d'EQI à référence complète. Il a une complexité de calcul pertinente si l'on considère ses applications diverses. Un code Matlab non optimisé prend environ 60 secondes afin de produire une estimation de qualité sur un processeur 1,8 GHz avec 2 Go de RAM sous Windows XP et Matlab R2008a pour une image de 512 x 768. La quantité de temps nécessaire pour l'apprentissage est négligeable, comme c'est le temps pris pour prédire la qualité par le classifieur qualifié / régresseurs par rapport à celui de l'extraction de caractéristiques. Les statistiques de corrélation spatiale occupent un morceau considérable de la durée de traitement.

La décomposition de la pyramide orientable dans cette version de DIIVINE est effectuée en utilisant la boîte à outils Matlab des auteurs de [64], sans utiliser de code MEX tel que recommandé. Etant donné qu'il existe un code C pour le même algorithme, il n'est pas faux de croire que le temps de calcul de cette section peut également être réduit considérablement.

Tableau 1.2 Le coefficient SROCC sur la base de données TID2008. Les algorithmes en Italique sont sans référence, et les autres sont des algorithmes à référence complète.

	JP2K	JPEG	WN	Gblur	All
PSNR	0.825	0.876	0.918	0.934	0.870
SSIM (SS)	0.963	0.935	0.817	0.960	0.902
DIIVINE	0.924	0.866	0.851	0.862	0.889

1.4.4 **En résumé**

Ce paragraphe a présenté un algorithme d'EQI sans référence et un algorithme intégré basé sur Statistiques scène naturelle, qui évalue la qualité d'une image sans nécessiter son image de référence, à travers une variété de distorsion catégories. L'algorithme DIIVINE utilise deux étapes : l'identification de la distorsion et l'évaluation de la qualité spécifique à une distorsion. Il fournit une mesure objective de la qualité perceptive, utilisant des paramètres statistiques extraits de la scène naturelle. Ces paramètres statistiques ont été détaillés, avec des références tirées des sciences de la vision et du traitement d'image, et il a été démontré que l'indice DIIVINE est bien corrélé avec la perception humaine de qualité. Une analyse approfondie de cet indice a ensuite été effectuée sur la base de données publique LIVE, et il a été montré (se référer au tableau 1.2) que la mesure proposée est plus performant que d'autres algorithmes d'EQI sans référence spécifiques à des types de distorsions donnés. Enfin, il a été démontré que les performances de DIIVINE sont indépendantes de la base de données choisie et peuvent facilement être étendues à d'autres types de distorsions que ceux qui ont été considérés ici, et une analyse de la complexité de calcul a été effectuée.

1.5 Méthode « Hybrides » : Métriques de Qualité basées sur le Système Visuel Humain (SVH)

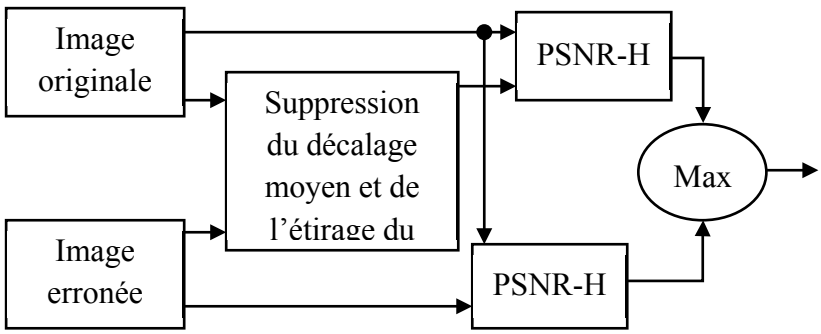
Plusieurs mesures objectives de qualité d'image existant dans l'état de l'art telles que l'Erreur quadratique moyenne (EQM), le PSNR, l'Erreur Absolue Moyenne (Mean Absolute Error (MAE)), etc, …, ne sont pas toujours corrélées avec l'évaluation subjective de la qualité. Deux images ayant des distorsions différentes et le même PSNR par rapport à l'image originale, peuvent par exemple donner lieu à une grande différence sur l'impact visuel. Dans ce travail [20], deux nouvelles métriques à référence complète sont proposées sur la base des propriétés du SVH. La première appelée PSNR-HVS est obtenu à partir d'un algorithme prenant en compte le SVH et la seconde est une version améliorée de l'Indice de Qualité Universel (Universal Quality Index (UQI)) dénotée UQI-HVS.

1.5.1 Description des métriques proposées

Un organigramme pour le calcul de la métrique proposée PSNR-HVS est présenté dans la figure 1.6. La suppression du décalage moyen et de l'étirage du contraste est effectuée en utilisant une fenêtre de balayage selon le procédé décrit dans [65]. Cette expérience utilise une taille de fenêtre de 64x64 pixels. Le PSNR obtenu s'écrit alors :

$$PSNR - H = 10 \log\left(\frac{255^2}{MSE_H}\right) \tag{1.2}$$

Dans cette équation, le MSE_H est calculé sur la base du SVH comme indiqué dans (1.3):



. Figure 1.6 Organigramme des étapes de calcul du PSNR-HVS

$$MSE_H = K \sum_{i=1}^{I-7} \sum_{j=1}^{J-7} \sum_{m=1}^{8} \sum_{n=1}^{8} \left(\left(X[m,n]_{i,j} - X[m,n]_{i,j}^{e}\right) T_c[m,n]\right)^2 \tag{1.3}$$

Où i,j désigne la taille de l'image, K = 1/[(I-7) (J-7) 64], Xij sont les coefficients DCT du bloc d'image 8x8 pour lesquels les coordonnées de l'angle supérieur gauche est égal à i et j. $X^{e}_{i,j}$sont les coefficients DCT du bloc correspondant dans l'image d'origine, et Tc est la matrice du facteur de correction.

En prenant en compte un SVH de haute sensibilité aux distorsions causées dans la gamme des basses fréquences, on a introduit la modification suivante sur l'UQI. Premièrement, une transformée en ondelettes discrète de premier ordre est appliquée à l'original et la déformation des images. Chaque image est alors divisée en 4 sous-bandes de fréquences: LL, HL, LH et HH. Ensuite, les valeurs de l'UQI sont calculées pour chaque sous-bande: ULL, UHL, ULH et UHH, respectivement. La valeur finale UQI-HVS est alors calculée comme suit:

$$UQI - HVS = 0.577U_{LL} + 0.1582U_{HL} + 0.1707U_{LH} + 0.0932U_{HH} \quad (1.4)$$

1.5.2 **Test des Images et Expérience subjective**

Dans cette expérience, la base de données d'images utilisée contient 44 images de test. Les images originales sont Lena et Barbara (niveaux de gris de 8 bits, la taille 512x512). Afin d'évaluer les performances de la métrique proposée, des dégradations conformes à celles causées par les trois niveaux de 7 distorsions différentes (bruit additif, bruit additif dû aux hautes fréquences, bruit impulsionnel en sel et poivre, flou, bruit de quantification, compression JPEG, compression JPEG2000) ont été produites par synthèse. Les niveaux de distorsion ont été choisis de telle sorte que la qualité visuelle des images de chaque groupe de test est aussi uniformément répartie que possible.

Pour identifier la relation entre la métrique objective proposée et la qualité perçue, une expérience psychophysique a été menée là où la valeur moyenne des nuisances sur des distorsions semblables a été mesurée. Trois groupes indépendants d'observateurs de l'Ukraine, de la Finlande et l'Italie, ont été invités à évaluer la qualité des images. Parmi ces sujets étaient des étudiants de premier cycle et des cycles supérieurs. Il leur a été demandé de porter leurs dispositifs habituels de correction de vision (lunettes ou lentilles). Le nombre total de sujets était de 56. Les images ont été affichées aux observateurs sur des écrans de haute qualité de 17 et 19 pouces de résolution 1152x864. L'expérience s'est déroulée en deux temps, pour une durée totale d'environ 20 minutes. Dans un premier temps, le tri des images est effectué par les observateurs à partir de la suite de test suivant leur qualité visuelle. Puis l'évaluation quantitative de la qualité d'une image donnée est déterminée par sa position dans l'ensemble disposé obtenu (1-21). Dans un second temps, les sujets devaient évaluer les nuisances des distorsions de ces images par des valeurs incluses dans l'intervalle [0,100]. Ici '100' correspond au cas où aucune distorsion n'a été détectée et '0 'au cas où les nuisances sont très ennuyeuses. Des instructions ont été données sous forme écrite aux sujets. Après lecture de ces instructions, une session de formation a été menée pour leurs indiquer les actions à mener. Les images ont été présentées dans un ordre aléatoire afin de minimiser l'effet contextuel. La validité des résultats du test subjectif a été vérifiée par criblage des observateurs selon à l'annexe 2 de l'UIT- R Rec. BT.500 [31]. Le MOS a été calculé en faisant la moyenne des sur l'ensemble des notes donnés par les observateurs. Plusieurs courbes d'ajustements ont été testées. La meilleure solution a été donnée par une courbe logarithmique (Figure 1.7).

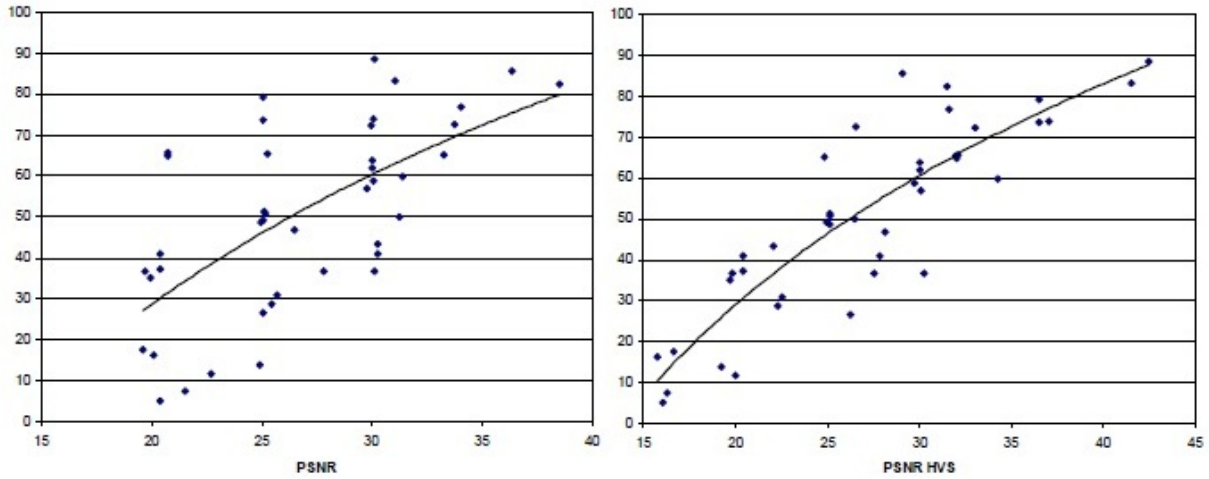


Figure 1.7 Comparaison des corrélation du PSNR et du PSNR-SVH avec le MOS [20].

La corrélation de Pearson entre les estimations quantitatives de la qualité visuelle de l'image obtenue pour deux parties de notre expérience était de 0,99. Ceci confirme indirectement son exactitude méthodologique.

1.5.3 **En résumé**

Dans ce travail, deux métriques sont créées pour l'évaluation de la qualité visuelle de l'image basée sur le PSNR (rapport signal de crête sur bruit) et le UQI (l'indice universel de qualité). Les résultats expérimentaux (Figure 1.7) montrent les avantages des mesures proposées par rapport à PSNR conventionnel, à l'UQI et au MSSIM.

1.6 Conclusion

L'évaluation de la qualité des vidéos ou images a connu beaucoup d'évolution ces dernières décennies. Les méthodes subjectives qui étaient les premières consistaient à faire évaluer une vidéo directement par un ensemble d'observateurs qui lui donnaient une note suivant la qualité de la vidéo. L'indice de qualité de la vidéo était ensuite calculé comme moyenne des différentes notes données par la vidéo. Cette méthode qui reste assez subjective et très relative aux sujets qui notent la vidéo atteint ses limites de point de vue pratique parce qu'elle n'est pas temps-réel et trop couteuse. Les chercheurs se sont donc ensuite orientés vers des méthodes objectives, c'est-à-dire qui ne seront pas relatifs à un jugement d'un groupe de personnes. Ces méthodes se classent en fonction de la nécessité ou non d'utilisation d'une image de référence pour l'évaluation de l'image. Ainsi, on distingue les méthodes d'EQI à référence entière, les méthodes d'EQI à référence partielle et les méthodes d'EQI sans référence. Des méthodes « hybrides » ont été développées par la validation de métriques objectives à l'aide d'expériences subjectives. C'est le cas des deux métriques présentées en section 1.5.

Cette thèse étant consacrée à la mise en œuvre d'un outil de contrôle de la qualité d'une vidéo, les prochains chapitres seront consacrés à la mesure de la qualité dans le décodage vidéo MPEG.

2 Evaluation de la Qualité d'une Vidéo par classification

2.1 Introduction

Avec l'augmentation des exigences pour les réseaux multimédias dans l'amélioration de la qualité de service (Quality of Service (QoS)) fidèlement au SVH, il y a eu de grands intérêts (cf. paragraphe 1.5) dans l'application de l'intelligence artificielle pour l'EQV. L'évaluation visuelle de la qualité d'une vidéo reste une étape fondamentale dans de nombreux systèmes de traitement vidéo. Toutefois, la seule manière de s'assurer que l'évaluation de qualité d'une vidéo est conforme à l'appréciation de la perception humaine est de la comparer aux méthodes subjectives basées sur le jugement humain (cf. section 1.3 p. 272). Jusqu'ici, les seules techniques procédant de la sorte ont été des métriques dites « hybrides » (cf. section 1.5 p. 35).

Le jugement humain ne se résume pas à une simple notation sur une échelle numérique mais opère plutôt (implicitement) par comparaison pour aboutir finalement à une classification de la qualité de la vidéo. La décision est alors prise en fonction de plusieurs facteurs déterminants, intrinsèques (bruit additif, flou, …) ou extrinsèques (complexité temporelle ou spatiale de la scène filmée, …). A l'opposé, notre idée est de procéder par une analyse objective de la qualité d'image et de proposer une méthode d'appréciation visuelle de la qualité d'une image conforme à la vision humaine.

Le chapitre est organisé comme suit: le paragraphe 2.2 décrit des travaux réalisés dans [8] sur l'EQV utilisant une combinaison de classifieurs ; le paragraphe 2.3 introduit l'outil mise en œuvre dans ce chapitre pour la mesure de la qualité de la vidéo utilisant la classification. Le paragraphe 2.4 décrit l'implémentation de l'OMQV basé sur la classification et présente les résultats de simulation. Enfin, le paragraphe 2.5 propose une conclusion.

2.2 Combinaison des classifieurs

L'idée d'utiliser la classification dans l'évaluation de la qualité d'une vidéo n'est pas complètement nouvelle. Une méthode de classification élaborée en marge de ce projet et basée sur les SVM est présentée en annexes du manuscrit. Des travaux ultérieurs [8] ont montré des avancées considérables dans le masquage et dans la classification. Les résultats de ces travaux sont commentés dans ce paragraphe.

Avec le développement de la connaissance du fonctionnement du SVH, la modélisation de certaines caractéristiques prenant en compte le SVH a été rajoutée dans le calcul de métriques « hybrides » telles que le PSNR-HVS. La prise en considération de ces métriques se fait dans l'espoir d'obtenir des résultats en accord avec le jugement humain sur la qualité visuelle des images.

Toutes ces métriques permettent d'obtenir un score de qualité permettant de procéder à un ordonnancement des niveaux de qualité. Néanmoins, l'un des inconvénients majeurs est l'utilisation de la métrique de Minkowski pour obtenir le score final, indépendamment du critère de décision utilisée (distance de Manhattan, distance euclidienne, etc.). Cependant en observant de plus près la façon dont le cerveau humain procède pour évaluer la qualité d'une image (avec ou sans connaissance de l'image originale). Ils ont pu constater que ce dernier ne donne pas directement de note à l'image, mais opère plutôt par comparaison pour finalement aboutir à une classification de la qualité de l'image. En effet, le score est affecté à l'image en fonction de plusieurs éléments déterminants, la démarche globale est donc plutôt une sorte d'approche multicritères. Cette approche ne semble pas opérer par un simple calcul d'une note moyenne représentant le niveau de distorsion causé par les défauts visualisés. Elle s'appuie sur un schéma de classification multi-classes. Ainsi, au lieu de procéder à un calcul de score représentant les critères mesurés, un vecteur multidimensionnel est créé et est directement utilisé par le schéma de classification. Les classes de qualité utilisées sont celles définies dans la norme ITU préconisées lors de la mesure de qualité par observations humaines.

2.2.1 Création et manipulation du vecteur de qualité

Cela consiste à définir un vecteur d'attributs en vue de l'apprentissage des classes de qualité des images. Parmi tous les attributs existants dans le domaine du traitement d'images, quelques critères ont été étudiés. Trois critères liés à la structure de l'image [65] ont été retenus. Ce sont : le critère de distorsion de luminance, le contraste et le critère de comparaison de structure. Deux critères colorimétriques permettant de mesurer la distorsion de chrominance, ainsi que la dispersion spatio colorimétrique. Avec en outre une mesure des effets de blocs, basée sur une décomposition de Fourier, une mesure d'effet de flou basée sur une méthode de détection de contours, ainsi que trois attributs (énergie, entropie et coefficient d'homogénéité).

La théorie de l'apprentissage statistique conduit à un algorithme appelé SVM (Support Vector Machines). Ce sont des vecteurs de support permettant de faire des estimations en matière de classification (et de régression). Une particularité de cette méthode est de produire une fonction de décision qui utilise un sous-ensemble de la base d'apprentissage. Les vecteurs de ce sous-ensemble sont appelés à Vecteurs de Support (VS).

Considérant une base d'apprentissage A = $\{(x_1, y_1), \dots, (x_k, y_k),\}$ composée de k couples avec xi ϵ Rn et yi ϵ $\{-1, +1\}$. L'algorithme des SVM projette les vecteurs xi dans un espace H à partir d'une fonction non linéaire φ : Rn $\rightarrow$ H. Ce qui permet alors de déterminer

l'hyperplan optimal de séparation des deux classes dans l'espace H. La frontière de séparation entre ces deux classes est matérialisé par hyperplan créé, noté (w, b). Etant donné un exemple x, sa classe y correspondante est donné par:

$$y = \text{sign}\,(w \,.\, \varphi(x) + b) \tag{2.1}$$

L'hyperplan est dit optimal si la distance qui le sépare des échantillons dont il est le plus proche est maximale. Cette distance est généralement appelée marge du classifieur. Les SVM étant des classifieurs binaires, ils effectuent la résolution d'un problème multi-classes en le représentant sous forme d'une combinaison de problèmes binaires [30]. Ainsi, dans cet application, cinq classes de niveau de qualité sont définies {Excellent, Bonne, MOyenne, MAuvaise, Très Mauvaise}.

Cependant, les SVM n'étant pas capables de calculer les probabilités a posteriori de classification, une mixture de gaussienne est calculée pour les 5 classifieurs. Considérant alors un classifieur $(C_i)_{i\in[1;:::;5]}$ et un exemple x_i à classer ; l'excentricité à la gaussienne est calculée, ce qui va indiquer la plausibilité que l'exemple xi soit correctement classifié. En considérant la confiance $p(C_i)$ accordée au classifieur Ci, sur le calcul d'une probabilité a posteriori, la confiance que l'on accorde à la classification d'un exemple x_i par le classifieur C_i est définie par la probabilité conditionnelle $p(x_i \mid C_i)$. La combinaison des résultats ainsi obtenus va alors résulter en une décision finale d'appartenance à l'une des cinq classes de qualité. La théorie de l'évidence, a été choisie pour cette application. Elle permet entre autre de traiter des informations incertaines, et de combiner des informations provenant de plusieurs sources.

L'ensemble des 5 classes de qualité est noté $\Omega = \{\omega_E, \omega_B, \omega_{MO}, \omega_{MA}, \omega_{TM}\}$, correspondant respectivement aux classes de qualité excellent, bonne, moyenne, mauvaise et très mauvaise. La théorie de l'évidence ne restreint pas ses mesures à cet ensemble Ω, mais s'étend sur l'ensemble 2^{Ω} des 2N sous-ensembles de Ω. Une fonction de masse m est ensuite défini, représentant la croyance que l'on accorde aux différents états du système, à un instant donné : m : $2^{\Omega} \rightarrow [0, 1]$ tel que $\sum_{A\subseteq\Omega} m(A) = 1$ et m(Ø) = 0

Où m(A) représente la croyance que l'on place dans l'hypothèse A ;

A représente soit un singletonω_E, soit une disjonction $\{\omega_E, \omega_{MO}\}$ de plusieurs hypothèses ; m(A) quantifie la croyance que la classe cherchée appartienne à A ;

Voici l'équation de combinaison de 2 fonctions de masse m_1 et m_2 selon la règle de Dempster :

$$m(A) = \frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K} \;,\; \begin{matrix}\forall A\epsilon\Omega \\ A\neq\emptyset\end{matrix} \tag{2.2}$$

Où K est le facteur de conflit entre les 2 sources et s'écrit $K = \sum_{B\cap C=\emptyset} m_1(B)m_2(C)$.

Le maximum de probabilité "pignistique" BetP(ω) est ensuite utilisé pour décider de l'élément le plus "probable" de Ω :

$$BetP(\omega) = \sum_{\omega\epsilon A\subseteq\Omega} \frac{m(A)}{|A|},\ \forall\,\omega\epsilon\Omega \tag{2.3}$$

Où $|A|$est le cardinal de A.

Le choix est alors porté sur un élément ω* de Ω auquel cette valeur est la plus grande :

$$\omega^* = Arg\{max_{\omega\epsilon\Omega}[BetP(\omega)]\} \tag{2.4}$$

La distance entre une vidéo x_i à classifier et le barycentre d'une classe de qualité ω_n a été formulée de deux différentes façons : d_E la distance euclidienne et d_{EP} la même distance pondérée par un critère de corrélation comparant la dispersion spatio-colorimétrique de deux nuages de couleurs.

2.2.2 Mesure des Performances

La méthode proposée a été testée sur 2 sous-ensembles d'images tirées de la base de donnée LIVE. L'un constitué d'images compressées par jpeg2k et l'autre constitué d'images bruitées par le bruit gaussien. Les DMOS $\overline{d_j}$ont été calculées à partir des MOS (score moyens d'opinions) de la manière suivante :

$$\overline{d_j} = \frac{1}{N}\sum_{i=1} N(r_{iref(j)} - r_{ij}) \tag{2.5}$$

où N représente le nombre d'opinions (d'observateurs) ;

$r_{iref(j)}$ la note donnée sur l'image de référence j par l'observateur i ;

et r_{ij} la note donnée sur l'image dégradée j par l'observateur i.

La Figure 2.1 représente les classifications faites pour l'ensemble des images constituant la base de données utilisée. La comparaison de cette méthode avec des méthodes de mesures de qualité existantes (MS-SSIM, VIF et VSNR) démontre ses performances.

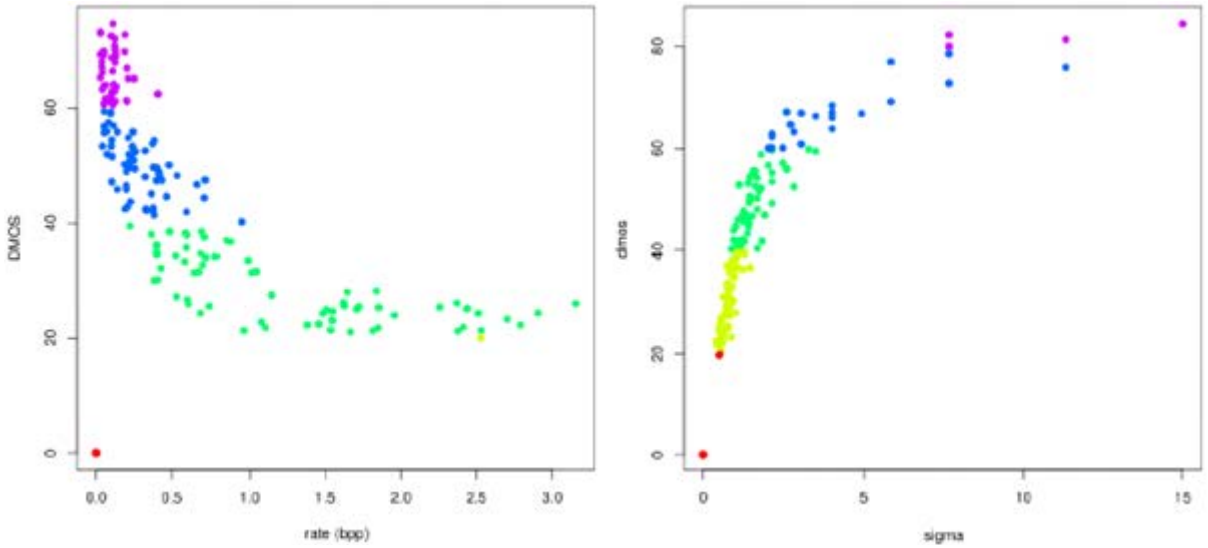


Figure 2.1 Couleurs associées aux 5 classes de qualité : violet-5 (E), bleu-4 (B), vert-3 (MO), jaune-2 (MA) et rouge-1 (TM) : à gauche : DMOS vs taux de compression JPEG2k ; à droite : DMOS vs σ du bruit gaussien.

La classification proposée est comparée aux métriques MS-SSIM, VIF et VSNR selon 3 critères : 1) le coefficient de Spearman (SROCC), 2) le coefficient de corrélation linéaire de Pearson (CC) et 3) l'erreur des moindres carrées (RMSE). Ces deux dernières mesures ont été réalisées après l'application d'une régression non linéaire [55].

Le tableau 2.1 présente le résultat de comparaison de la méthode proposée avec différentes métriques de qualité connues de l'état de l'art. Ces résultats prouvent que les performances de l'approche proposée rivalisent avec les performances des trois autres métriques.

Tableau 2.1 Comparaison des Performances des différentes méthodes [8].

Base JPEG2K					
	méthode avec		MS-SSIM	VIF	VSNR
	d_E	d_{EP}			
SROCC	0.9614	0.9649	0.9647	0.9720	0.9406
CC	0.9609	0.9712	0.9708	0.9789	0.9476
RMSE	6.0199	5.9151	5.9899	5.0925	5.6143
Base du bruit gaussien					
	méthode avec		MS-SSIM	VIF	VSNR
	d_E	d_{EP}			
SROCC	0.9340	0.9699	0.9519	0.9706	0.9507
CC	0.9401	0.9677	0.9487	0.9762	0.9501
RMSE	5.9776	5.7435	5.8225	4.9905	5.5443

La méthode proposée est une approche originale de mesure de la qualité des images. Elle peut se targuer de rivaliser avec les métriques très connus du domaine de l'Evaluation de la Qualité d'une Vidéo (EQV) sur un même rang et est même meilleure que ces dernières sur la base du taux de reconnaissance. Les performances en termes de coefficients de corrélation, qui s'apparentent à ceux obtenus avec les trois autres métriques, sont sans équivoque dus aux fonctions noyaux utilisées qui ne sont pas optimisées par rapport aux données du problème. Cette approche peut être appliquée quelque soient les défauts que l'on souhaite classifier : dans ce cas, il suffit d'adapter le classifieur et le processus d'apprentissage et de test aux caractéristiques du problème.

2.3 OMQV basé sur la classification

Le reste du chapitre présente une méthodologie de prédiction du niveau de qualité d'une vidéo. La proposition consiste en l'utilisation d'un algorithme de classification simple et générique qui permet d'estimer le niveau de qualité d'une vidéo donnée.

2.3.1 Généralités sur l'OMQV à base de la classification

L'objectif principal est de classer objectivement le niveau de qualité d'une vidéo selon l'échelle continue de l'UIT-T à 5 niveaux [32], conformément au jugement humain sur la qualité de la vidéo. Le défi est de créer un outil de mesure objective de la qualité d'une vidéo travaillant directement à partir des mesures de qualité de vidéo disponibles, pour la mise en œuvre d'un système de classification entre 5 classes de qualité vidéo considérés (excellent, bon, moyenne, mauvais et très mauvais) et la moyenne d'opinion en scores (MOS) donnés par des humains. Des résultats prometteurs sont obtenus en utilisant la classification k-NN entraînée sur un ensemble de données d'une expérience subjective combiné avec des indicateurs fondamentaux mesurables, à savoir le taux de perte de paquets, le rapport signal de crête sur bruit ainsi que les index spatiaux et les indices temporels. Une analyse statistique permettra ensuite de comparer les performances de cette solution avec des ensembles de données obtenus par évaluation subjective de l'homme.

Le choix a été porté sur la classification k-NN parmi divers critère de classification (Naive bayes, Perceptron, machines de vecteurs, etc…). Elle s'adapte mieux aux besoins du problème d'évaluation du niveau de qualité d'une vidéo connaissant des paramètres descriptifs de cette vidéo. Pour le calcul de la distance au plus proche voisin, nous avons opté pour la distance euclidienne à cause de sa simplicité.

Ce problème peut être comparé à celui de l'analyse typologique, où il s'agit d'analyser un seul ensemble de données et décider si et comment les observations dans l'ensemble de données peuvent être divisés en groupes. Dans une certaine terminologie, en particulier celle de machines d'apprentissage, le problème de classification est classé comme un type d'apprentissage supervisé.

2.3.2 Conception de l'OMQV basée sur la classification

Dans la reconnaissance de formes, l'algorithme des k plus proches voisins k-NN (k Nearest Neighbors) est un procédé de classification d'objets à partir d'exemples de formation les plus proches dans l'espace des attributs. La classification k- NN est un type d'apprentissage basé sur les instances, ou l'apprentissage paresseux où la fonction est seulement approchée localement et tout calcul est reporté jusqu'à ce classement.

Les voisins sont pris à partir d'un ensemble d'objets pour lesquels la classification correcte (ou, dans le cas de la régression, la valeur de la propriété) est connue. Cela peut être considéré comme l'ensemble de la formation pour l'algorithme, si aucune mesure explicite de formation n'est nécessaire. L'algorithme k-NN des k plus proches voisins est sensible à la structure locale de ces données. En effet la règle d'estimation du voisin le plus proche calcule la limite de décision d'une manière implicite. Il est également possible de calculer la limite de décision de façon explicite, et de le faire de manière efficace de sorte que la complexité de calcul soit une fonction de la complexité de frontière.

Les échantillons devant être mémorisés durant l'apprentissage sont des vecteurs dans un espace de caractéristiques multidimensionnel, chacun avec une étiquette de classe. La phase d'apprentissage de l'algorithme consiste essentiellement en la mémorisation les vecteurs de caractéristiques et des étiquettes de classe des échantillons d'apprentissage.

Dans la phase de classification, k est une constante définie par l'utilisateur, et un point non marqué est classé par l'assimilation à l'échantillon qui est le plus fréquent parmi les k échantillons de formation les plus proches de ce point.

Habituellement, la distance euclidienne est utilisée comme mesure de la distance, mais ce n'est applicable que pour les variables continues. Dans des cas tels que la classification de texte, des métriques telles que le chevauchement (ou distance de Hamming) peuvent être utilisés. Souvent, la précision de la classification "k"-NN peut être améliorée de manière significative si la métrique de distance est enseignée avec des algorithmes spécialisés, comme par exemple l'algorithme du plus proche voisin avec "grande marge" ou l'analyse des composantes de voisinage.

Pour la conception de l'OMQV basée sur la classification, les échantillons à classifier sont des vidéos décodées. Les paramètres d'entrée du classifieur sont des valeurs de métriques de qualité vidéo. La phase apprentissage concerne 80% de la base de données.

A- Base de données utilisée

Ce travail utilise la base de données de [17] et [18]. La base de données utilisée ainsi que les paramètres vidéo sont décrits dans cette section. Les séquences vidéo considérées sont de niveaux de qualité visuelle différents. Elles ont été obtenues par simulation d'un réseau artificiel propice aux erreurs de perte de paquets. Six séquences vidéo nommément Foreman, Hall, Mobile, Mother, News et Paris, cinq différents taux de perte de paquets PLR [0,1 %, 0,4 %, 1%, 3%, 5 %, 10%] et deux réalisations de Canaux ont été considérées pour la mise en place de cette base de données. Toutes les Séquences originales de sont disponibles au format progressif brut au débit de 30 images par seconde.

Cette base de donnée a été choisie parce qu'elle contient des séquences ayant différents niveaux de complexité spatiale et temporelle (voir Figure 2.2):

- L'information spatiale perceptuelle (SI)

L'information spatiale [2] perceptuelle, SI (Spatial Index), est basée sur le filtre de Sobel. Chaque image vidéo (plan de luminance) à l'instant n (Fn) est d'abord filtrée par le filtre de Sobel [Sobel (Fn)]. L'écart-type sur les pixels (std_{space}) dans chaque trame filtrée par le filtre Sobel est ensuite calculé. Cette opération est répétée pour chaque trame dans la séquence vidéo, ce qui résulte en une série temporelle de données spatiales de la scène. La valeur maximale de cette série chronologique (max_{time}) est choisie pour représenter l'information spatiale SI de la scène. Ce processus peut être représenté sous forme d'équation:

$$SI = max_{time} \{std_{space} [Sobel(Fn)]\} \quad (2.6)$$

- L'information temporelle perceptuelle (TI)

L'information perceptuelle temporelle TI (Temporal Index) [2], est basée sur la fonction de différence de mouvement, M_n (i, j), qui est la différence entre les valeurs de pixel (sur le plan de luminance) au même emplacement dans l'espace, mais à des moments ou des trames successives. M_n (i, j) comme une fonction de l'instant (n) est définie de la manière suivante:

$$M_n (i, j) = F_n (i, j) - F_{n-1}(i, j) \quad (2.7)$$

Ici F_n (i, j) est le pixel à la ligne i et de la colonne j de la n-ième trame dans le temps. La mesure de l'information temporelle TI, est calculée comme étant le maximum sur le temps (max_{time}) de l'écart type dans l'espace (std_{space}) de M_n (i, j) sur tout i et j.

$$TI = max_{time} \{std_{space} [M_n (i, j)]\} \quad (2.8)$$

La grande mobilité dans des trames adjacentes se traduira par des valeurs élevées de TI.

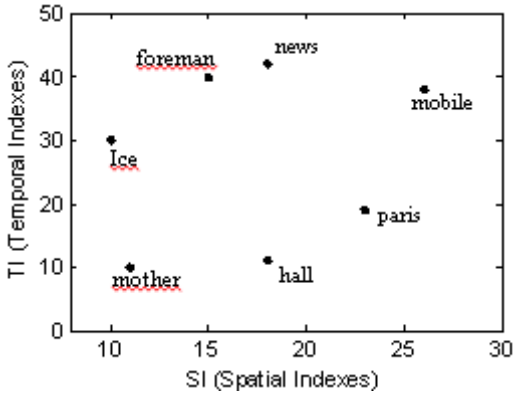


Figure 2.2 Valeurs de SI et TI des vidéos de la base de données.

Il est précisé que ces séquences ont été choisies de sorte de pouvoir couvrir différents niveaux de Complexité spatiale et Temporelle. Une autre raison du choix de cette base de données est la mise en disposition des moyennes des opinions d'observateurs (MOS) établies et fournies avec la base de données

Le Logiciel "transmitter_simulator" présenté en [17] simule la transmission du flot de données H.264/AVC sur un canal de transmission susceptible de contenir des erreurs. Les trois principaux éléments dans la conception de ce canal artificiel sont les suivants:

• Possibilité de traiter le flot de données en paquets aussi bien avec la normalisation de l'annexe B qu'avec celle du protocole de transfert en Temps Réel RTP (*real-time transfert protocol*).

• Différentes Modalités de corruption du flot des données: tous les paquets, tous les paquets contenant uniquement les tranches codées en intra et tous les paquets contenant des tranches codées en inter.

• Possibilité de générer différentes réalisations de canal à partir d'un seul fichier de configuration d'erreurs.

Les fichiers de signature des erreurs utilisés pour dégrader le flot de données :

Chaque motif d'erreur consiste en une séquence de 10000 Caractères ASCII de '0 ' et de '1'. Un '0 ' en i-ème position du fichier de configuration d'erreur signifie que la transmission du i-ème paquet codé a été reçu avec succès. Inversement, la présence d'un '1' en i-ème position dans le fichier de configuration d'erreur signifie que le i-ième paquet codé a été perdu au

cours de sa transmission. La séquence de zéros et de uns à l'intérieur de chaque fichier de configuration d'erreurs a été générée suivant le modèle de Gilbert à deux états dont les paramètres peuvent être ajustés pour obtenir le PLR désiré et la longueur moyenne de rafale. Une rafale d'erreurs dans un canal est défi comme une séquence contiguë de deux ou plusieurs '1'.

Les paramètres du modèle ont été configurés de manière à obtenir une longueur moyenne de rafale égale à 3 paquets. Pour chaque valeur du PLR, différentes réalisations ont été obtenues en commençant la lecture du motif d'erreur à partir d'un point aléatoire. Chaque séquence vidéo a 298 trames (images); Une réalisation est définie par un PLR donné et un offset, c-à-d au premier '0' sur la signature d'erreur correspondant au premier paquet RTP reçu ; chaque trame contient un nombre fixe de 18 tranches ; Qui sont des paquets NALU où chaque paquet contient une tranche codée, qui contient 22 MB (Macrobloc). En outre, chaque MB dans la norme de codec vidéo H.264/AVC (utilisée en [18]) contient 16x16 pixels, soit 256 pixels.

B- Les métriques choisis pour l'évaluation de la qualité des multimédias

A partir d'une vidéo donnée, on peut extraire différents paramètres qui permettent de déterminer le niveau de qualité de la vidéo. Voici les quelques mesures sélectionnées dans cet ouvrage:

- **Le taux de perte de paquets PLR**

La perte de paquets survient lorsqu'un ou plusieurs paquets de données transitant sur un réseau ne parviennent pas à atteindre leur destination. La perte de paquets est considérée comme l'un des principaux types d'erreurs rencontrés dans les communications numériques. Les deux autres étant erreurs sur les bits erronés et les paquets parasites causés par le bruit. Le taux de perte de paquets peut déterminer immédiatement si la séquence vidéo est l'objet d'une bonne ou mauvaise réception. En effet, plus le PLR (*packets loss rate*) est élevé (trop de paquets perdus pendant la transmission), plus il y a de la discontinuité dans les pixels constituant la trame et donc plus la qualité de l'image ou de la vidéo est mauvaise (voir Figure 2.4 et Figure 2.3). Le taux de perte au cours de la transmission est donc un indice du niveau de qualité du canal de la communication. Donc, le PLR est une bonne mesure à prendre en compte lors de l'évaluation de la qualité de la vidéo reçue.

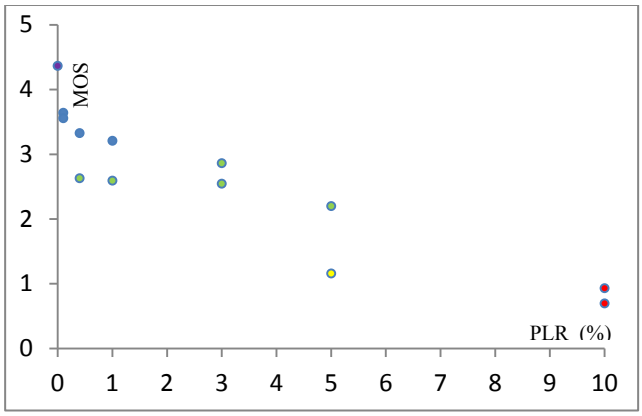


Figure 2.3 Dépendance de la qualité d'une vidéo sur le taux de perte de paquets.

Compte tenu de l'élaboration de la base de données décrite ci-dessus établissant une valeur de PLR pour chaque vidéo, ce PLR sera simplement extrait de la vidéo et constituera la métrique PLR de la vidéo.

D'un point de vue exclusivement statique, la métrique PLR de taux de perte de paquets dans l'évaluation de la qualité d'une vidéo ainsi définie peut être transposée à une notion de pertes de paquets dans le cadre d'évaluation de la qualité d'une image. C'est dans ce cadre que nous avons considéré le PLR dans l'application de l'OMQV lors de l'évaluation des images plus loin dans le dernier chapitre du manuscrit.

Figure 2.4 Une frame de la vidéo "foreman" :de gche à dte : originale, avec 0,4% de PLR et avec 5% PLR.

Une des plus anciennes métriques sera également prise en compte parmi les attributs (paramètres) sélectionnés pour l'outil de mesure de la qualité d'une vidéo à base de classification, à savoir le PSNR.

- **Le pic du rapport signal sur bruit PSNR**

Le pic du rapport signal sur bruit, expression souvent abrégé PSNR, est un terme d'ingénierie pour le rapport entre la puissance maximale possible d'un signal et la puissance de bruit qui affecte la fidélité de sa représentation. Parce que de nombreux signaux ont une très

large gamme dynamique, le PSNR est généralement exprimé en termes d'échelle logarithmique des décibels.

Le PSNR est le plus souvent utilisé comme un indicateur de la qualité de la reconstruction des codecs de compression avec perte (par exemple, pour la compression d'images). Les données d'origine dans ce cas représentent le signal, et le bruit est l'erreur introduite par compression. Lorsque l'on compare les codecs de compression, il est utilisé comme une approximation de la perception humaine de la qualité de la reconstruction, donc dans certains cas, une reconstruction peut apparaître comme proche de l'original que l'autre, même si elle a un PSNR inférieur (un PSNR supérieur devrait normalement indiquer que la reconstruction est de meilleure qualité). Il faut être extrêmement prudent avec le domaine de validité de cette paramètre car il est concluant et valide uniquement quand il est utilisé pour comparer les résultats d'un même codec (ou type de codec) et d'un même contenu. Pour pallier à cet inconvénient du PSNR, des chercheurs ont élaboré d'autres métriques liées au SNR et plus représentatives de la perception visuelle telles que le VSNR [7].

Le PSNR est plus simple à définir par le biais de l'erreur quadratique moyenne MSE (mean square error) qui pour deux images monochromes m × n I et K, considérant l'une des 2 images comme donnée : une approximation du bruit présent dans l'autre image est déterminée par:

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{2.9}$$

Le PSNR est défini par:

$$PSNR = 10.\log_{10}\left(\frac{{MAX_I}^2}{MSE}\right) \tag{2.10}$$

$$= 20.\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) \tag{2.11}$$

Ici, MAX_I est la valeur maximale possible de pixel de l'image. Lorsque les pixels sont représentés sur 8 bits par échantillon, elle est de 255. Pour les images en couleur avec trois valeurs RGB par pixel, la définition du PSNR est le même, sauf que le MSE est la somme de toutes les différences de valeur au carré divisé par la taille de l'image et par trois.

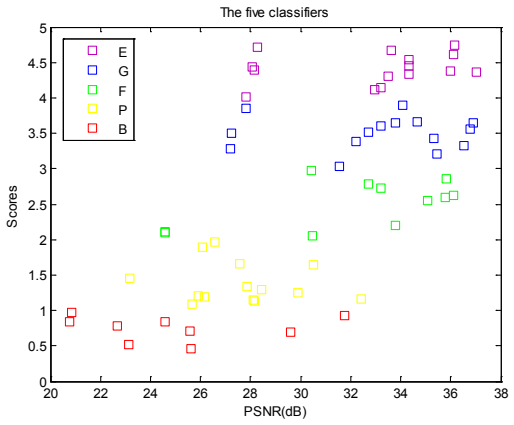


Figure 2.5 Classes de qualité : MOS vs. PSNR pour la phase validation.

Les valeurs typiques pour le PSNR dans les images avec perte et la compression vidéo sont entre 30 et 50 dB, où les plus hautes valeurs sont les meilleurs. Les valeurs acceptables pour la perte de qualité de transmission sans fil sont considérées comme étant d'environ 20 dB à 25 dB. Lorsque les deux images sont identiques, le MSE sera égal à zéro sauf que pour cette valeur, le PSNR est indéfini.

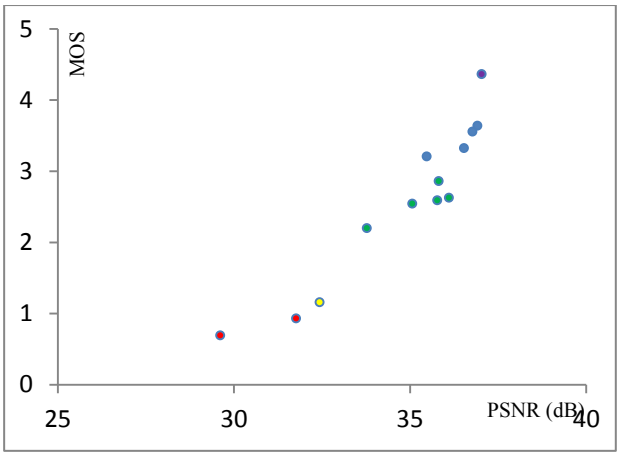


Figure 2.6 Qualité de la vidéo vs. PSNR pour la séquence Vidéo "foreman".

Le PSNR est dit être la métrique objective à référence complète la plus utilisée, avec un faible coût de calcul, des significations physiques, et il est mathématiquement facile à implémenter à des fins d'optimisation. Cependant, il a été largement critiqué pour ne pas être en bonne corrélation avec la mesure perceptuelle de la qualité.

Les résultats des expériences décrites dans [17] et [18] montrent que l'évaluation de la qualité des vidéos est plus pertinente avec le PSNR pour un certain type de distorsions ainsi que pour différentes fréquences spatiale et les distributions de couleur du bruit additif. Cela confirme un phénomène typique pour le SVH: sensibilité élevée au bruit spatialement corrélées et sensibilité faible au bruit dans les composants de couleur. Les images dégradées par le bruit spatialement corrélées ont une qualité visuelle plus mauvaise que les images dégradées par du bruit gaussien présentant un PSNR plus bas (moins de 6 dB). La Figure 2.6 montre la dépendance de la qualité d'une vidéo (mesurée par le MOS) sur le PSNR.

- **L'information spatiale perceptuelle (SI)**

Nous voyons dans la Figure 2.2 et avons mentionné en début de ce paragraphe que les paramètres SI et TI constituent une façon unique d'identifier une scène donnée. Comme mentionné à l'introduction générale, pour des raisons d'homogénéité entre le traitement d'images et celui des vidéos dans ce projet nous nous sommes exclusivement intéressé à l'aspect statique et avons exclus les effets liée à la complexité temporelle de la vidéo. Ainsi, la métrique TI donnant l'information temporelle de la vidéo ne pouvant pas retrouver son sens dans le cadre d'une image, nous ne l'avons pas choisi parmi nos métriques d'évaluation de la qualité des vidéos.

Lorsque deux images sont identiques, l'erreur quadratique moyenne MSE (*mean squared error*) entre les deux images ou des images vidéo sera égale à zéro. La Figure 2.6 montre les valeurs MOS en fonction du PSNR des vidéos pour les séquences vidéo d'origine "foreman" et ses 12 vidéos générées à partir des 12 différents niveaux de perte de paquets. Cette

dépendance de la qualité des vidéos sur le PSNR est monotone, tout comme la dépendance de la qualité des vidéos sur le taux de perte de paquets.

- **La métrique blurMetric, pour la mesure du Flou**

L'objectif du projet étant de réutiliser l'OMQV dans l'algorithme de correction des erreurs sur une image, il était nécessaire d'ajouter une métrique aux 3 premières considérées dans l'EQI. Nous avons jugé bon d'ajouter une métrique servant à la mesure du flou, qui est un artefact dominant dans le décodage MPEG. La métrique utilisée dans ce projet pour la mesure du flou a été établie par F. Creté et Al. [15] en partant du fait que l'ajout de flou dans une image varie suivant le niveau de flou déjà présent dans cette image.

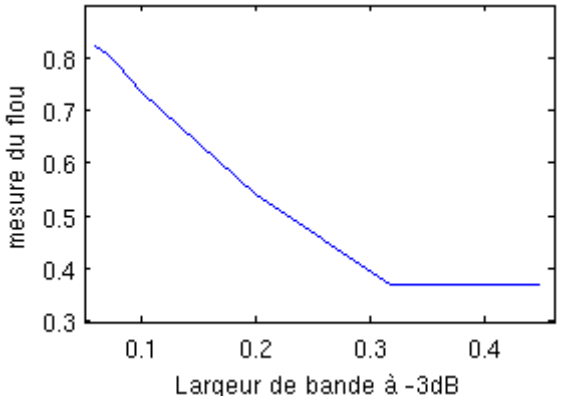


Figure 2.7 Variation du niveau de flou en fonction du paramètre de floutage BT. Où BT est le produit entre B (la bande passante unilatérale) et T (la période d'échantillonage.

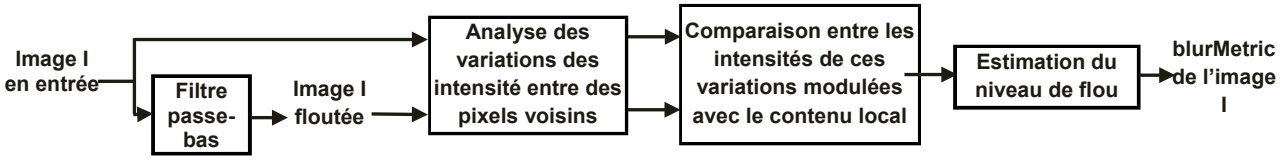


Figure 2.8 Schéma bloc simplifié de l'estimation de la métrique flou.

L'algorithme de calcul de cette métrique quantifie le flou sur une image en la floutant d'avantages à l'aide d'un filtre passe bas et en comparant les différences entre les pixels voisins, avant et après le passage du filtre passe bas. Cette métrique n'est toutefois pas applicable à certaines images ayant une texture homogène et une très faible complexité spatiale c-à-d un SI très faible. La Figure 2.7 indique que la métrique choisie assure une monotonie avec le paramètre BT de variation du niveau de flou.

La Figure 2.8 présente le schéma bloc simplifié du principe d'estimation de la métrique flou utilisée. Le côté pratique et la simplicité de conception de cette métrique sont les raisons qui

expliquent le choix de son utilisation dans le présent projet. Le calcul de cette métrique est expliqué en Annexe, au paragraphe A.3.2 b).

Une fois les paramètres de classification et la mise en œuvre des classes définis, place maintenant à l'implémentation de l'outil d'évaluation basé sur la classification.

2.4 Implémentation de l'OMQV

L'algorithme de classification proposée crée d'abord une fonction de prédiction des scores vidéo à partir d'un apprentissage de 80% des séquences vidéo de la base de données choisis aléatoirement et leurs MOS correspondants. La fonction de prédiction est basée sur les k voisins les plus proches selon le critère de la distance euclidienne, en prenant k = 1 pour simplifier l'algorithme. Dans un deuxième temps, les vidéos ou les images sont classées en fonction de leurs scores prédits selon les 5 niveaux de qualité recommandés par l'UIT illustré dans la table 6.1. En outre, une couleur est attribuée à chacune des cinq classes de qualité vidéo (Figure 2.5). Le reste des séquences vidéo (soit 20% des séquences vidéo de bases de données) ont été utilisées pour valider (10% de la BD) puis tester (10% de la BD) l'outil de classification créé.

2.4.1 Apprentissage du classifieur

Détermination des classes de qualité

Le système doit pouvoir classe de nouvelles vidéos (ou de nouvelles images) (lors de la phase de validation par exemple) en se basant sur les exemples mémorisés pendant la phase d'apprentissage, qui soient les plus proches dans l'espace des 4 différents paramètres. L'algorithme k-NN est un type d'apprentissage basé sur les instances, (ou l'apprentissage "paresseux") où la fonction est uniquement estimée localement. Les images voisines sont prises à partir d'un ensemble d'images pour lesquelles la classification correcte est connue. L'algorithme k-NN des k-plus proches voisin est sensible à la structure locale des données. En effet, la règle des k voisins les plus proches calcule la limite de décision d'une manière implicite. Etant donné le nombre de voisins ramené à 1, l'équation de calcul du plus proche voisin est alors simplifiée.

La base de données contient 6x(1+12) = 78 séquences vidéo dont 80% (62 séquences vidéo) sont utilisées pour l'apprentissage. Lors de la phase d'apprentissage, le système prend en entrée une matrice 62x4 constituées de 62 séquences vidéo. Les vidéos (et leurs MOS respectifs) sont soumises une à une au système d'apprentissage et chaque vidéo est représentée par un vecteur vi = [PLR_i, $PSNR_i$, SI_i, $blurMetric_i$] (et son MOS respectif par MOS_i) représentant les caractéristiques (paramètres) de la vidéo (le MOS correspond à la note de qualité de la vidéo).

Soit E = {(v_i, MOS_i), i = 1,…,62} l'ensemble des données soumis à l'apprentissage. Pour une vidéo donnée v_0, la détermination du plus proche voisin v_i suivant la distance euclidienne est donnée en (2.12) :

$$d(v_0, v_i) = \sqrt{\begin{matrix}(PLR_0 - PLR_i)^2 + (PSNR_0 - PSNR_i)^2 + (SI_0 - SI_i)^2 \\ +(blurMetric_0 - blurMetric_i)^2\end{matrix}} \quad (2.12)$$

Classification de la vidéo

Pour attribuer une classe à cette vidéo v_0, le système va d'abord classer les 62 vidéos d'apprentissage dans leurs classes de qualité suivant leur MOS.

Une fois le score S0 d'une vidéo v0 à classer déterminé, un algorithme de classification basé sur les 5 niveaux de qualité indiqués dans la Figure 1.2 se charge de classer la vidéo suivant les 5 classes de qualité:

- Si 5≥ S0 ≥ 4 la vidéo est dans la classe Excellente
- Si 4> S0 ≥ 3 la vidéo est dans la classe Bonne
- Si 3> S0 ≥ 2 la vidéo est dans la classe Moyenne
- Si 2> S0 ≥ 1 la vidéo est dans la classe Mauvaise
- Si 1> S0 ≥ 0 la vidéo est dans la classe Très mauvaise

Cet algorithme simplifie le classement de vidéo une fois l'estimation du score MOS0 effectué.

Une fois les 5 classes constituées, un calcul des barycentres des classes permet de déterminées les vidéos moyennes virtuelles $\bar{v}_j$ pour chacune des 5 classes.

$$\overline{\boldsymbol{vJ}} = \left(\frac{\sum_{j-1<MOSi\leq j} \boldsymbol{xi}}{\boldsymbol{Card(Classej)}} \right) \textbf{ ; j=1 à 5}$$

La classe de v_0 sera celle de la vidéo moyenne $\overline{\boldsymbol{v_{jmoy}}}$ de la classe j située à d_{min} de v_0 telle que:

$$d_{min} = min_j(d(v_0, \overline{\boldsymbol{v_J}})) \quad (2.13)$$

La phase de validation que nous décrirons dans la section suivante donne les résultats de simulation de ce système sous Matlab.

2.4.2 Validation du classifieur

Les deux étapes décrites ci-dessus ont été implémentées en un algorithme en code Matlab et simulées. Les résultats obtenus sur les 24 séquences réservées à la validation sont présentés dans cette section.

La Figure 2.10 montre les résultats obtenus par l'outil de classification lors de la phase d'apprentissage et de validation. Les résultats de la phase de test (Figure 2.11, Figure 2.12 et

Figure 2.13) montrent que l'outil effectue un très bon classement dans les classes "Très mauvaise", "Moyenne" et "Bonne". Ces performances ont été établies sur des séquences vidéo en Format CIF de la base de données utilisée. On peut noter que l'OMQV basé sur la classification donne de très bons résultats (MSE <0,05) pour les trois classes de niveau de qualité satisfaisante (excellente, bonne et moyenne), ce qui signifie que le classifieur pourra toujours bien distinguer des vidéos de qualité visuelle satisfaisante du reste.

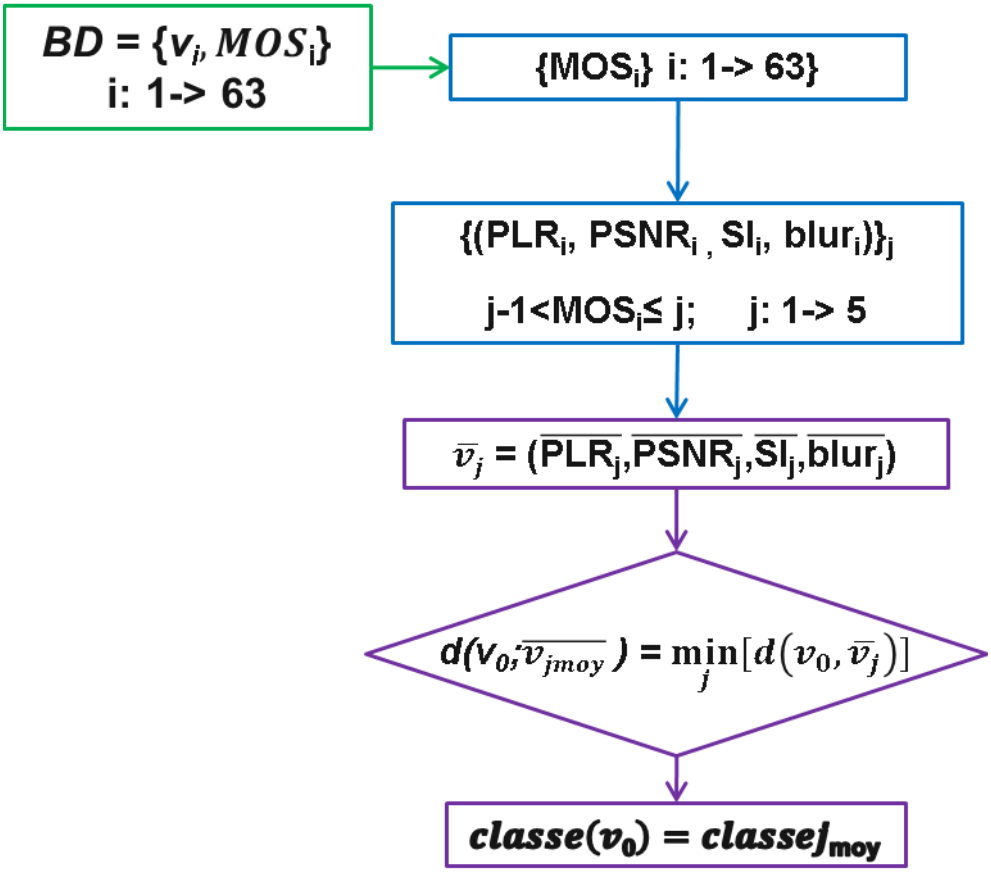


Figure 2.9 : Algorithme de classification d'une vidéo v_0.

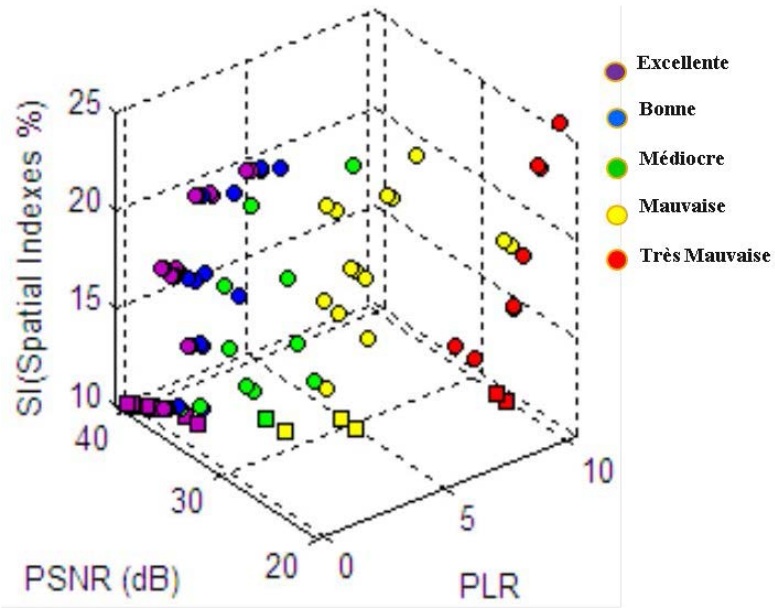


Figure 2.10 Vidéos d'apprentissage (rond) et validation (carré): (PLR, PSNR, SI). La couleur violette est pour la classe Excellent, bleue pour la classe Bon, verte pour la classe moyenne, jaune pour la classe Mauvais et rouge pour la classe Très mauvais.

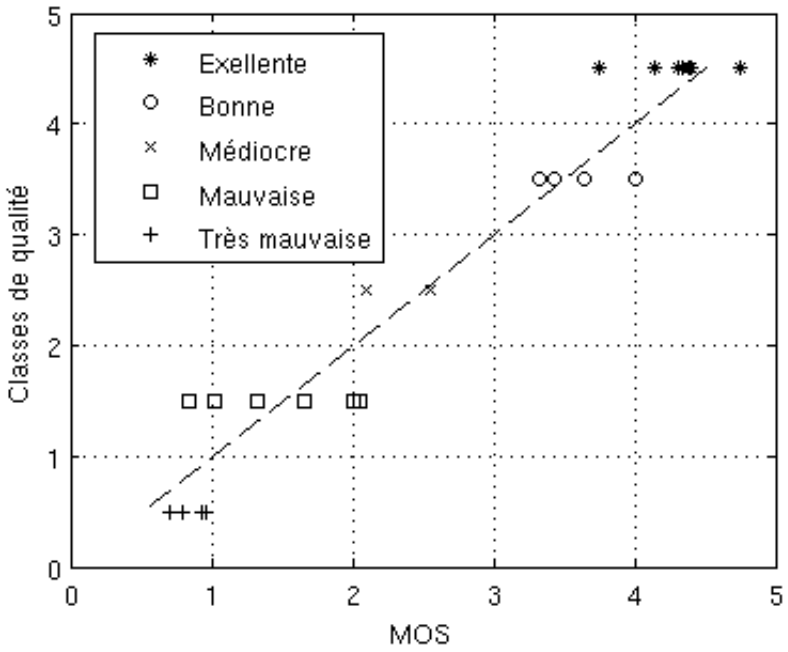


Figure 2.11 Résultats de classification pour la phase de test.

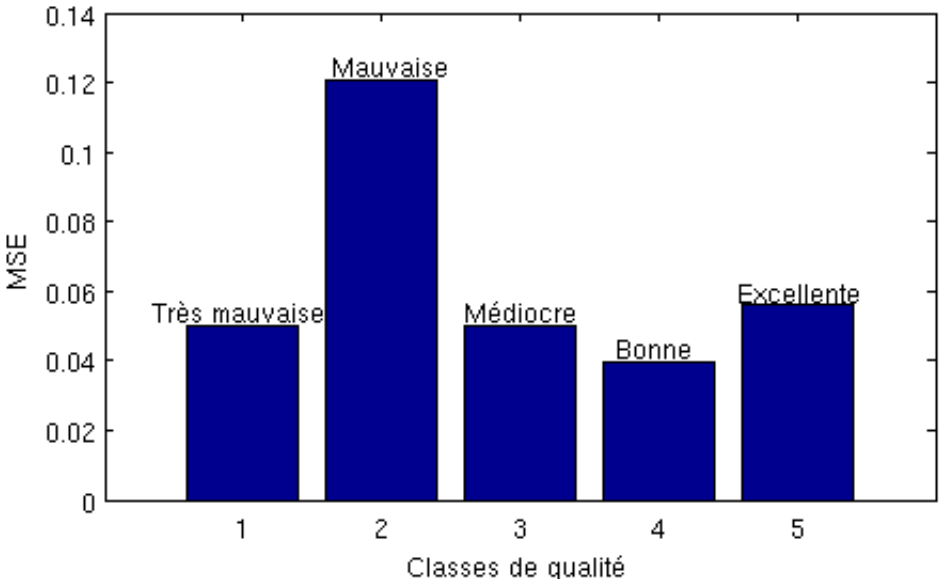


Figure 2.12 Erreurs quadratiques moyennes obtenu sur la classification des vidéos de validation.

Le taux de réussite de l'outil de classification sur chacune des classes a été calculé pour déterminer les performances du classifieur à retrouver les niveaux de qualité d'une vidéo par rapport aux MOS donnés par les observateurs humains. Pour une classe de qualité C donnée, le taux de réussite sera égal au nombre de vidéos v classifiées dans C et dont le score S appartient au même niveau de qualité (Figure 1.2) que son MOS, sur le nombre total de vidéos de la même classe C. La Figure 2.13 montre le taux de réussite avec lequel l'OMQV a classé les images durant la phase de validation. Ce taux de réussite démontre la concordance avec le MOS.

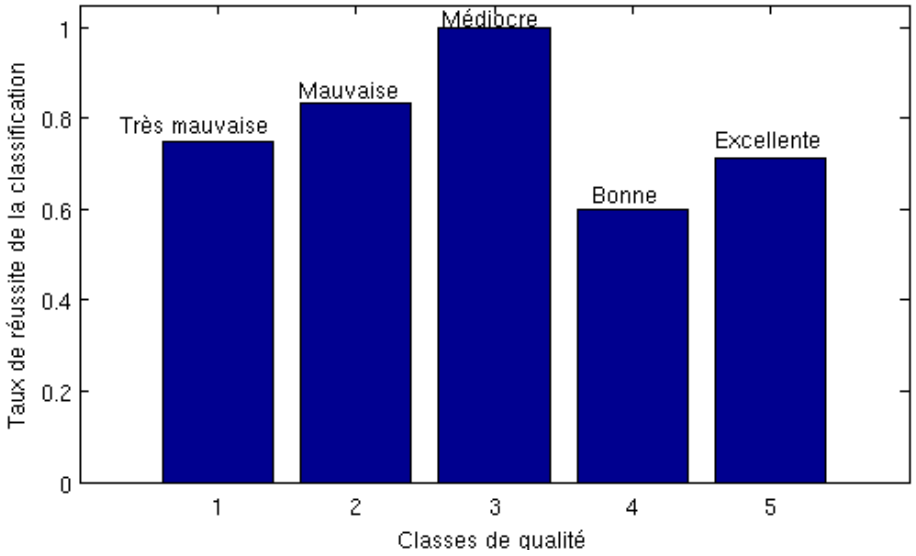


Figure 2.13 Taux de réussite du classifieur sur les 5 classes de qualité vidéo.

La concordance entre le classifieur de qualité proposé et les moyennes d'opinion des scores (MOS) permet de confirmer que l'outil de mesure de la qualité par classification proposé ici donne une évaluation du niveau de qualité d'une vidéo en corrélation avec l'intelligence humaine. En effet, les valeurs heuristiques et nominales du MOS données en section 6.2.1 sont en parfaite adéquation avec la classification proposée par l'OMQV basé sur la classification.

Test sur l'indépendance par rapport au contenu

Un test de performances a été effectué sur le format 4CIF et a donné des résultats plutôt médiocres (Figure 2.14). En effet, la classification n'a pas été performante selon les valeurs de MOS observés. Elle manque de précision notamment pour la classe de qualité vidéo "Bonne". Quatre sur sept (soit 57,14 %) des vidéos classées comme étant de qualité "Excellente" ne sont pas correctement classées. Une d'elle (soit 14,28 %) devrait d'après son MOS (compris entre 2 et 3) correspondre à la classe de qualité "Moyenne" et les autres (soit 42,86 %) devraient être classées comme "Bonne".

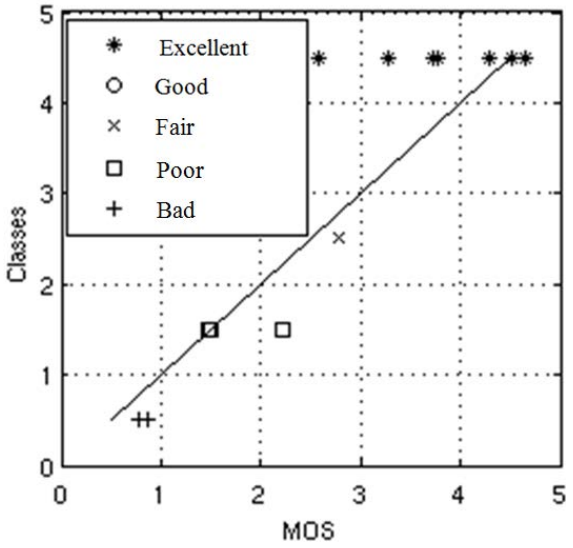


Figure 2.14 Classification obtenue pour les séquences vidéo de format 4CIF.

La classification des vidéos de format 4CIF a obtenue : 100% de réussite pour la classe Bad ; 100 % pour la classe Poor ; 33 % pour la classe Fair ; 0 % pour la classe Good et 100% pour la classe Excellent. En tout seules 8 sur 13 (soit 61,53 %) des vidéos ont été correctement classées. Ces résultats indiquent la nécessité d'inclure tout nouveau format vidéo à l'apprentissage de l'OMQV par classification pour espérer atteindre des résultats satisfaisants dans l'utilisation du classifieur sur la mesure de qualité de vidéos de ce nouveau format.

2.5 Conclusion

Dans ce chapitre, nous avons présenté un outil de mesure de la qualité d'une vidéo basé sur la classification avec le critère du plus proche voisin selon la distance euclidienne, que nous avons développé et implémenté. L'OMQV basé sur le classification permet une évaluation objective de la qualité d'une vidéo en la classifiant dans l'échelle de qualité de 5 niveaux (Très mauvais, Mauvais, Moyenne, Bon, Excellent). On obtient un classement correct dans l'ensemble, avec un pourcentage de réussite de 78% sur l'ensemble des classes de qualité. Des performances satisfaisantes ont été obtenues avec un MSE moyen de 0,063. Par ailleurs, l'outil conçu offre une solution suffisamment simple et générique pouvant être utilisée soit pour le contrôle de la qualité dans un réseau multimédia, soit pour l'évaluation de la qualité d'un décodage vidéo numérique de type MPEG. Bien que l'application à tout nouveau format de vidéo nécessite un apprentissage de ce nouveau format, l'OMQV basé sur la classification permet néanmoins d'ouvrir la voie à de nouvelles techniques de mesure de qualité, plus proche de ce qui est réalisé par le système visuel humain, autrement que par une simple classification.

3 Evaluation de la Qualité d'une Vidéo par Réseaux de Neurones Artificiels

3.1 Introduction

En plus d'être conforme au jugement humain sur la qualité d'une vidéo, un outil d'EQV pourrait noter une vidéo en lui donnant un score numérique de qualité. Une évaluation de la qualité d'une vidéo qui restitue un score numérique comme l'expérience [17-18] peut s'avérer indispensable dans certaines applications. Une telle évaluation remplacerait alors complètement l'intervention humaine dans les expériences d'EQV. Nous avons ainsi pensé aux méthodes d'intelligence artificielle pour apporter cette particularité à notre OMQV.

Le souci étant de préserver une grande cohérence entre le jugement humain (MOS) et la note attribuée à l'image, une alternative serait de développer un OMQV basé sur les réseaux de neurones artificiels (RNA). D'où l'idée [36] de modéliser un RNA capable de remplacer le cerveau humain en fournissant une note de qualité visuelle similaire au MOS. Les réseaux de neurones artificiels sont généralement utilisés dans des problèmes de nature décisionnelle tels que la prédiction de l'évolution des cours dans les marchés boursiers, l'identification d'empruntes digitales, ….

Ce chapitre est organisé comme suit: le paragraphe 3.2 décrit des travaux réalisés en [9]. Le paragraphe 3.3 introduit l'outil d'intelligence artificielle mise en œuvre pour la mesure de la qualité d'une vidéo; le paragraphe 3.4 décrit l'implémentation de l'OMQV basé sur les réseaux de neurones artificiels et présente les résultats de simulation. Enfin, le paragraphe 3.5 conclut le chapitre.

3.2 Mesures de qualité d'image utilisant l'approche RNA

Des travaux antérieurs ([9] et [27]) ont montré des avancées considérables dans l'évaluation de la qualité d'une vidéo utilisant les RNAs. Les résultats des travaux référencés en [9] sont commentés dans ce paragraphe.

Dans [9], Chetouani et al. considèrent que la meilleure façon d'évaluer la qualité de l'image est d'abord d'identifier pour chaque distorsion la plus appropriée, en termes de corrélation avec les MOS. Ensuite d'essayer de combiner toutes les métriques pour construire une métrique globale et multi-distorsions. Cette stratégie a donc été adoptée pour développer une métrique de qualité d'image sans référence. Ce travail est porté sur les artefacts les plus communs et les plus ennuyeux qui sont causées par le processus de compression à savoir les effets de bloc, le flou et les suroscillations.

Les effets de bloc apparaissent dans la méthode de compression de bloc en tant que limites horizontales et verticales dans l'image. Ils sont dus au fait que les blocs sont transformés et quantifiés de façon indépendante. Certaines métriques sans référence ont été développées pour mesurer ou supprimer cet artefact. Dans [66], une méthode d'estimation aveugle de l'effet de bloc dans le domaine des fréquences est proposée. Dans [6], [14], les artefacts de blocs sont mesurés dans le domaine DCT. Dans [33], un algorithme itératif est appliqué pour réduire l'effet de bloc dans le domaine des ondelettes. Dans [11], l'effet de bloc est retiré en tenant compte de certaines propriétés du SVH.

Le flou affecte les contours et les détails dans l'image. Ceci est causé par la nature du processus de compression. En fait, les détails et les contours dans l'image correspondent aux hautes fréquences de l'image et généralement, le processus de compression affecte en premier lieu les hautes fréquences. Cet artefact est également largement étudié. Dans [42] une méthode d'estimation de flou à référence partielle est proposée. Une métrique d'EQV à référence partielle est également proposée dans [43], où le détecteur de bord est appliqué à l'image originale. Dans [63] une méthode d'estimation du flou basée sur une combinaison des hautes fréquences dans le domaine ondelettes est proposée. Dans [10], une nouvelle estimation de flou récente basée sur certains modèles de SVH est également proposée. Un autre élément intéressant fondé sur le SVH propose une métrique de mesure du flou sans référence a été proposée dans [25].

L'effet de suroscillations est aussi un des artefacts les plus ennuyeux et apparaît dans l'image comme du bruit sur les contours et il est plus visible lorsque les contours sont à proximité de régions homogènes. Cet artefact est causé par le procédé de quantification. Ce phénomène est également appelé Gibbs phénomène [22]. Dans [43], une métrique sans référence basée sur une estimation de l'oscillation autour des contours dans le domaine spatial est proposée. Dans [54], l'estimation des suroscillations est effectuée dans le domaine d'ondelettes.

Ce travail ([9]) propose une métrique d'EQI utilisant un modèle de fusion basé sur le réseau de neurones pour mieux profiter de certaines métriques efficaces. Cette méthode vise à estimer les distorsions les plus ennuyeux et connues à savoir, les suroscillations, l'effet de bloc et le flou. L'idée de base est d'utiliser les mesures quantitatives de ces dégradations en tant qu'entrées d'un Réseau Neuronal Artificiel (RNA). La sortie de ce réseau de neurones est une valeur unique correspondant au niveau de qualité.

L'apport essentiel de l'outil que proposé est non seulement de fournir une EQI conforme au jugement humain, mais capable aussi de restituer un score de qualité pouvant se substituer au MOS donné par les expériences subjectives.

3.2.1 La méthode proposée

Dans cette étude, un RNA est utilisé pour mesurer la qualité d'une image sans référence. L'organigramme de la méthode proposée est illustré sur la Figure 3.1. Les caractéristiques de l'image sont d'abord extraites. Il s'agit de trois métriques sans référence. Ces métriques sont utilisées pour mesurer respectivement les effets de bloc, le flou et les suroscillations. Ensuite, les caractéristiques obtenues sont utilisées comme entrées pour le RNA.

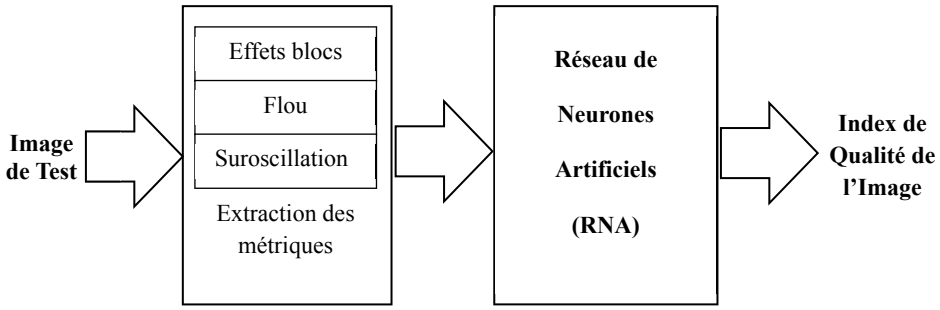


Figure 3.1 Organigramme de la méthode proposée.

La méthode utilise les métriques proposées par [66], [10] et [54] respectivement pour l'effet de bloc, le flou et les suroscillations. L'effet de bloc est mesuré sur l'analyse de la visibilité des blocs sur les bords de l'image. La première étape consiste à calculer la différence de l'image entre des rangées adjacentes (la même opération est appliquée horizontalement au niveau des bords). Ensuite, la somme de toutes ces mesures est effectuée. Enfin, après l'application d'une détection de zéro à la différence d'images obtenue, l'indice global de qualité est obtenu par pondération de ces valeurs. La métrique du flou est élaborée en ajoutant du flou à une image de test. Ensuite, l'impact de ce flou ajouté est mesuré à l'aide d'une analyse radiale réalisée dans l'espace dans le domaine fréquentiel comme indiqué en [4]. La métrique des suroscillations est basée sur la transformée en ondelettes et les scènes statistiques naturelles. Le calcul de l'indice de qualité est déduit de la transformée en ondelettes des coefficients d'un modèle statistique, après un processus de formation.

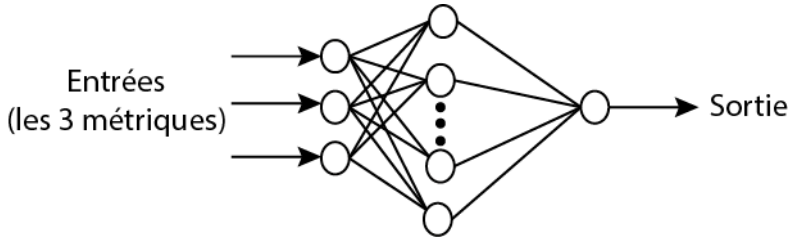


Figure 3.2 Modèle du RNA.

Une fois les caractéristiques extraites, ces fonctionnalités sont combinées en un RNA pour obtenir un indice de qualité. Le RNA utilisée ici et présenté à la Figure 3.2 est du type multicouches. Les scores subjectifs et objectifs extraits sont d'abord mis à l'échelle de la gamme [-1, +1] où un score de -1 désigne la meilleure qualité et un score de +1 dénote la plus mauvaise qualité. Cela est dû à la fourchette de fonction d'activation. Le nombre de neurones d'entrée est égal au nombre des éléments, ici trois. En sortie, il n'y a qu'un seul neurone. Pour la couche cachée, pour ne pas augmenter la complexité, le nombre de couches cachées a été fixé à un. La fonction sigmoïde est utilisée comme fonction d'activation dans les couches

cachées et de sortie. A noter que cette fonction est utilisée par le Groupe VQEG (Visual Quality Experts Group), d'experts de la qualité visuelle [29] pour décrire la relation entre scores objectifs et scores subjectifs.

3.2.2 L'apprentissage et le test

La base de données LIVE [56] a été utilisée pour l'entraînement et le test du RNA. Cette base de données fournit la différence moyenne des scores d'opinion DMOS et est décrit par les auteurs comme suit: «Les observateurs ont été invités à fournir leur perception de la qualité sur une échelle linéaire continu qui était divisé en cinq intervalles égaux marqués d'adjectifs "très mauvais", "mauvais", "moyenne", "bon" et "excellent". Environ 20 à 29 observateurs humains ont évalué chaque image. Chaque type de distorsion a été évalué par différents sujets dans des expériences différentes utilisant le même équipement et les mêmes conditions de visualisation ». Il est rappelé qu'un DMOS de valeur zéro correspond à une haute qualité d'image, alors qu'un DMOS de haute valeur reflète une mauvaise qualité de l'image. La base de données utilisée contient différents types de dégradations. Ici, la performance de la méthode a été évaluée en utilisant uniquement les images de flou gaussien.

La base de données choisie a été divisée en 2 sous-ensembles:

- Apprentissage: utilisé dans le processus d'apprentissage et contenant environ 100 images par dégradation.
- Test: utilisé uniquement pour tester l'efficacité du projet Procédé. Cet ensemble contient environ 50 images par dégradation.

La phase d'apprentissage a été effectuée en utilisant le procédé de rétropropagation. Les meilleurs résultats sont obtenus avec 19 neurones dans la couche intermédiaire. Enfin, la structure du RNA-MLP est la suivante:

- Entrée: 3 (mesures NR)
- couche cachée: 1 avec 19 neurones
- Sortie: 1 (DMOS cibles mis à l'échelle à l'intervalle [-1, +1])

3.2.3 Résultats expérimentaux

Après le processus d'apprentissage, l'efficacité de la méthode a été testée en utilisant la suite de test établie. Le tableau 3.1 et le tableau 3.2 montrent la corrélation obtenue par cette méthode pour chaque distorsion et pour toutes les distorsions, respectivement.

Comme l'indique le tableau 3.2, les coefficients corrélations de Pearson et Spearman obtenue pour les trois dégradations sont respectivement égaux à 0,89 et 0,86.

Tableau 3.1 Corrélation obtenue pour chaque type artefact.

Résultats de la méthode	
Dégradation	Coefficient de Pearson
Mesure des effets de bloc	0,9570
Mesure du flou	0,8462
Mesure des suroscillations	0,9289

Tableau 3.2 Corrélation obtenue pour tous les types d'artefact.

Type de corrélation	Valeur du coefficient
Pearson	0,89964
Spearman	0,86214

3.2.4 Résumé et synthèse

Un nouvel indice de la qualité de l'image sans référence est proposé par cette expérience. Cette méthode est basée sur l'utilisation par un RNA de métriques sans référence. La combinaison de ces métriques grâce au RNA donne de bons résultats en termes de corrélation avec l'évaluation subjective exprimée dans la base de données de qualité d'image LIVE.

Dans notre approche, nous nous sommes inspiré de la manière dont les observateurs évaluent la qualité d'une vidéo (ou d'une image) et nous remplaçons le jugement humain dans une telle expérience par un OMQV qui modélise une évaluation de la qualité des vidéos (ou des images) par un RNA. La différence ici est que c'est la vidéo (ou l'image) en question qui est considérée en entrée. Bien évidemment le réseau de neurones artificiels traitera des paramètres extraits de la vidéo (ou de l'image) décrivant la vidéo (ou l'image) en elle-même qui vont constituer les entrées du RNA.

3.3 OMQV basé sur les RNA

Le reste du chapitre présente une technique de prédiction du score de qualité d'une vidéo basée sur les réseaux de neurones artificiels (RNA). Le choix des RNA parmi les solutions proposées dans l'intelligence artificielle est motivé dans la prochaine section.

3.3.1 Concepts de base

Les travaux de recherche décrits en [19] utilisent des tests subjectifs pour prouver la corrélation entre l'activité spatiale et la compression des images. En [20], des expériences subjectives donnant lieu à des MOS ont été utilisés pour prouver l'efficacité de nouvelles métriques prenant en compte la perception visuelle humaine.

L'objectif étant de fournir un index sur la qualité d'une image avec la plus grande corrélation possible avec le SVH, il serait intéressant de considérer un outil de mesure de qualité basé sur une intelligence artificiel simulant le jugement humain sur la qualité d'une image. Parmi les outils de l'intelligence artificielle, les RNAs restent celui qui s'adapte le mieux à un problème de reconnaissance d'objets ou de formes pouvant se substituer (ou presque) au cerveau humain. D'où la considération des RNAs comme importante alternative à la problématique qui se pose.

Un RNA est composé de neurones artificiels interconnectés (un système programmé qui simule les propriétés des neurones biologiques). Les RNAs pourraient servir soit pour comprendre les réseaux neuronaux biologiques, soit pour résoudre des problèmes d'intelligence artificielle sans nécessairement créer un modèle du système biologique réel. Le système nerveux biologique réel est très complexe et inclus certains éléments qui pourraient paraitre d'aucune utilité si on se référait aux RNAs.

Un perceptron (la terminaison d'un neurone) est essentiellement un "classifieur linéaire" qui classifie des données $x \in \mathbb{R}^n$ spécifiés par les paramètres $w \in \mathbb{R}^n$, $b \in \mathbb{R}$ et la fonction de sortie $f = w'x + b$. Ses paramètres sont réglés par une loi ad-hoc semblable à la décroissance du gradient d'une grandeur stochastique. Comme le produit scalaire est un opérateur linéaire dans l'espace des entrées, le perceptron peut correctement classifier uniquement des données pour lesquelles différentes classes sont linéairement séparés dans l'espace des entrées tandis qu'il échoue souvent complètement pour des données non-séparables.

3.3.2 Conception de l'OMQV basé sur les RNAs

L'apport majeur de l'OMQV utilisant un RNA est la substitution du jugement humain (à travers les MOS) dans l'évaluation de la qualité d'une image, ou dans la mesure "différentielle" de la valeur ajoutée à une image par un algorithme de correction d'erreurs. Cet apport scientifique contribue de ce fait au contrôle de la qualité des décodages numériques, notamment de la famille des décodeurs MPEG.

Les RNAs [48] ont connus d'énormes succès dans des applications diverses dans presque tous les domaines de la technologie et de la science. Ils sont notamment utilisés dans les applications de reconnaissance de formes ou d'objets qui permettent la modélisation de "l'intelligence humaine".

L'implémentation d'un modèle fonctionnel artificiel de neurone biologique [34], nécessite de prendre en compte trois éléments de base, nommément la synapse, la dendrite et l'axone. La synapse du neurone biologique est l'élément qui fait l'interconnexion entre deux couches d'un réseau de neurones et donne la force de la connexion. Dans un neurone artificiel, les synapses sont représentées comme des poids. Un poids négatif reflète une connexion inhibitoire, tandis qu'une connexion positive reflète une connexion excitatrice. Les dendrites sont représentées par une combinaison linéaire de toutes les synapses (qui lui sont connectées). Ainsi, toutes les entrées sont combinées et la somme est biaisée par les poids. Enfin, une fonction d'activation (l'axone) contrôle l'amplitude de la sortie.

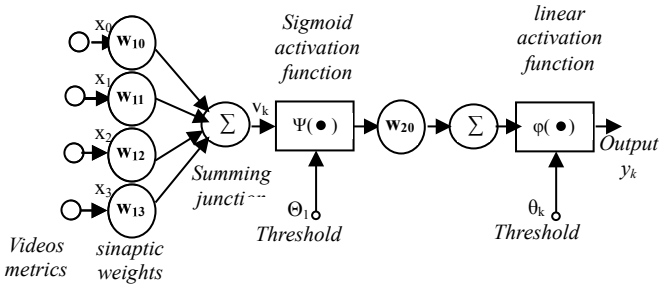


Figure 3.3 Architecture du réseau de neurones artificiels.

La fonction d'estimation des MOS basée sur les RNA sera donc assimilable à une combinaison d'un ensemble de poids calculés à partir des métriques d'EQV extraits de la vidéo. A partir du modèle représenté dans la figure 3.3, l'activité du neurone peut être décrite mathématiquement comme suit:

$$v_k = \sum_{j=1}^{k} \omega_{kj} x_j \tag{3.1}$$

L'axone ou la sortie du neurone, y_k, serait donc le résultat d'une fonction d'activation de valeur v_k comme indiqué dans l'équation (3.1).

La base de données et les métriques de qualité décrites dans la section 2.3.2 à la page 44 restent valables dans ce chapitre. Nous utiliserons donc la base de données disponible en [18] et les métriques PLR, PSNR, SI et blurMetric.

3.4 Implémentation de l'OMQV basé sur les RNAs

3.4.1 Modélisation et apprentissage du RNA

Dans la configuration du RNA prédestiné à l'estimation des scores de qualité des vidéos, les fonctions d'activation signum (Figure 3.2 a) et linéaire (Figure 3.2 c) ont respectivement été choisies pour les sorties ψ de la première couche et φ de la seconde couche du RNA. Ces fonctions d'activation ont été choisies au niveau de la sortie de chaque couche du RNA de façon à obtenir la précision d'une erreur MSE sueil de 0,05 avec les MOS.

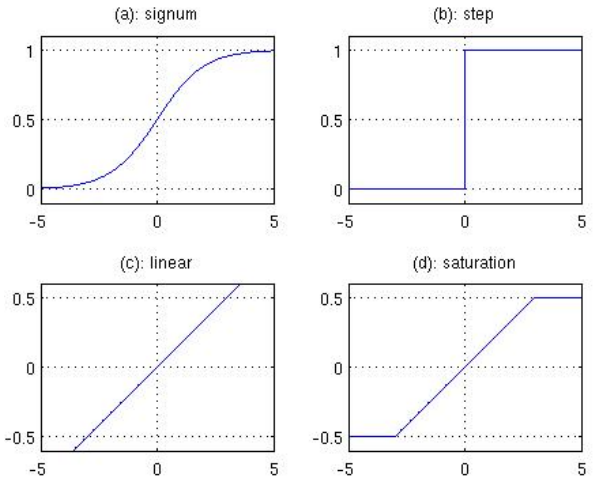


Figure 3.4 Fonctions d'activation usuelles.

On distingue trois types d'apprentissage : L'apprentissage supervisée, où le RNA est entraîné en lui fournissant à la fois les entrées et les sorties correspondantes attendues ; l'apprentissage non supervisée où le RNA est entraîné à fournir tout seul des sorties convergeant vers la sortie idéale attendue ; et l'apprentissage forcée intermédiaire entre ces

deux premières formes d'apprentissage. Le système d'apprentissage utilisé pour la conception du RNA a été l'apprentissage supervisé. Les caractéristiques d'entrée (pour l'apprentissage du RNA) correspondent à des paramètres des vidéos et les caractéristiques cibles correspondent à leurs valeurs respectives de MOS, tous deux prises de la base de données disponible en [17].

L'OMQV basé sur les RNAs a été conçu avec une architecture (Figure 3.1) de 4 entrées correspondant aux 4 paramètres (PLR, PSNR, SI et blurMetric), 1 sortie correspondant au MOS estimé pour la vidéo à évaluer, 10% des 13x6 = 78 échantillons ont été prises pour effectuer des tests, 10% pour la validation et le reste (80%) pour l'entraînement du RNA.

Pour obtenir des résultats expérimentaux les plus proches de la mesure abstraite et « universelle » (Figure 1.1) il est nécessaire d'obtenir le plus grand nombre possible de jugements humains (représentés par les MOS) sur la qualité des vidéos ou des images. Pour cela, la moyenne des deux bases de données de MOS des études faites respectivement par EPFL et Polimi ([17-18]) a été considérée durant toutes les phases (entraînement, tests et validation) de conception.

3.4.2 Validation et résultats de simulation

Cette deuxième méthode d'évaluation a été simulé (voir figure 3.5) et simulée sur Matlab. Le RNA réalisé a 2 couches avec 4 neurones pour la première couche et 1 neurone pour la seconde.

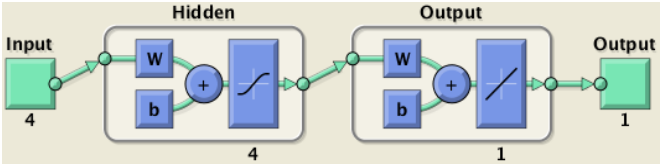


Figure 3.5 Architecture du RNA implémenté pour l'OMQV.

La Figure 3.4 présente le modèle Matlab Simulink du RNA implémenté. Une fois entraîné, il est simple à l'utilisation et nécessite juste d'introduire en entrée des valeurs des 4 métriques de la séquence vidéo à évaluer.

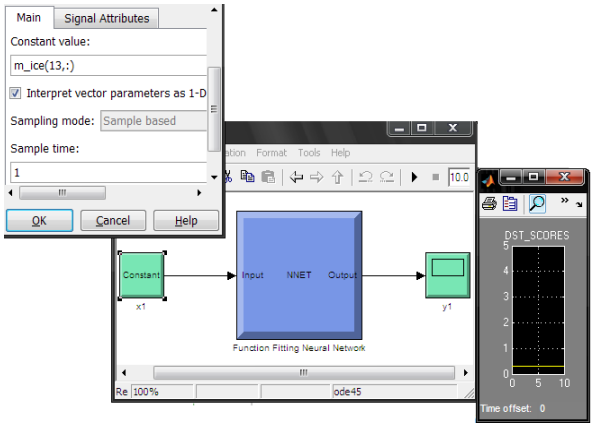


Figure 3.6 Capture d'écran de l'évaluation d'une séquence vidéo par le RNA.

Voici les étapes à effectuer pour évaluer la vidéo par l'OMQV base sur les RNAs implémenté:

Le vecteur des métriques (correspondant ici au vecteur note m_ice(13,:) dans la Figure 3.4) est donné dans le tableau 3.3. Il correspond aux 4 métriques de qualité vidéo PLR, PSNR; SI et blurMetric de la vidéo dont l'évaluation est présentée sur la figure 3.6.

Une vidéo de test (la séquence ice de format 4CIF [18]) a été évaluée, suivant des valeurs de PLR différentes. Considérons la séquence ice à un PLR de 10% (voir le tableau 3.3 pour les entrées et Figure 3.6 pour l'évaluation). L'OMQV a estimé le score (correspondant ici à la valeur numérique indiquée par le trait jaune dans la Figure 3.4, soit un score de 0.66 sur 5) de qualité de la vidéo dont les métriques sont données en entrée au RNA sur l'échelle de 0 à 5 présentée à la Figure 1.2.

Tableau 3.3 Vecteur d'entrée (métriques) de la séquence ice avec un PLR de 10%.

PLR	PSNR	SI	blurMetric
10.0000	26.3400	10.0000	0.3788

Le MSE obtenu à l'itération 19 est environ de 5.10-2 qui est erreur négligeable, preuve d'une grande précision. La valeur du coefficient de détermination SCC = 0,98 obtenue et la régression indiquée par la Figure 3.5 montre une étroite corrélation entre les MOS des expériences [17-18] et les MOS estimés par l'OMQV basé sur les RNAs.

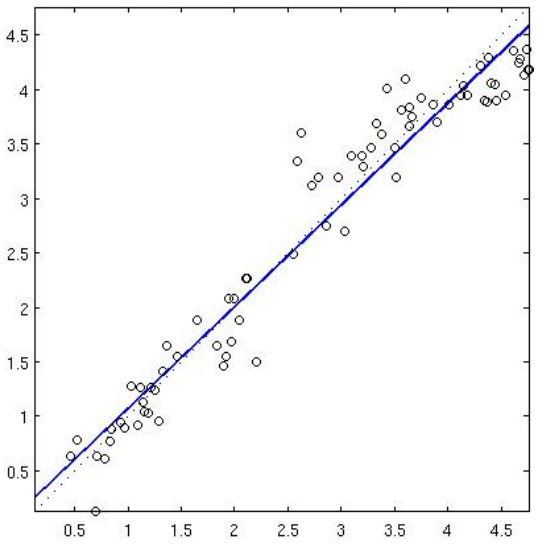


Figure 3.7 MOS vs. scores estimés par RNA à la fin de la phase d'apprentissage.

La Figure 3.6 montre une corrélation satisfaisante entre les scores estimés par l'OMQV basé sur les RNAs et les MOS. Ce résultat est aussi vérifié par les paramètres statistiques obtenus:

Coefficient de Spearman: 0,97137

Erreur moyenne entre les scores estimés et les MOS: 0,13.

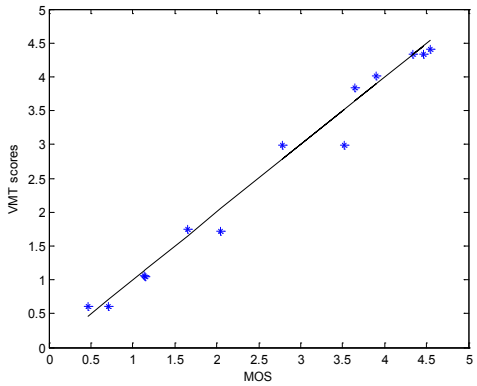


Figure 3.8 Corrélation entre MOS estimés et MOS pour la phase de test.

Le système a estimé le score de la vidéo ice sur l'échelle de qualité de la vidéo en continu. Le résultat obtenu est représenté sur la figure 3.9. Ce résultat montre bien une corrélation avec les MOS, bien qu'on note une linéarité déviée par rapport à la première bissectrice. Cette expérience prouve qu'on obtiendrait de bons résultats si on reprenait l'expérience avec des vidéos de format 4CIF durant la phase d'apprentissage. Donc les résultats obtenus sont garantis pour tout type de format de vidéo, mais à condition d'en inclure le format durant la phase d'apprentissage.

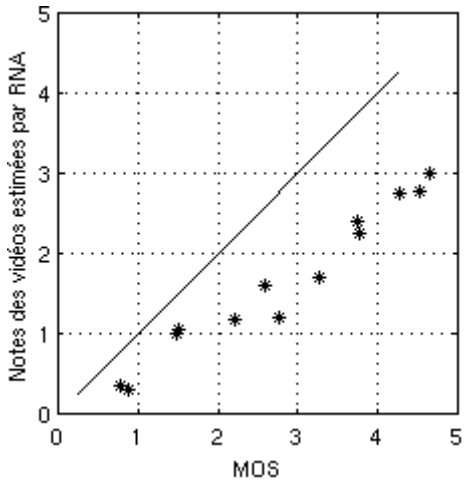


Figure 3.9 MOS vs. MOS estimés par le RNAs, pour une vidéo au format 4CIF.

3.5 Conclusion

Un outil de mesure de la qualité d'une vidéo (OMQV) basé sur les RNAs a été développé dans ce chapitre. L'outil estime un score numérique correspondant à l'évaluation de la qualité visuelle de la vidéo. La méthode utilise un apprentissage supervisé du RNA en prenant en entrée les paramètres de qualité de la vidéo PLR, PSNR, SI et blurMetric et génère en sortie un score sur une échelle de 0 à 5 qui (on l'a montré) est en corrélation avec la notion du jugement humain sur la qualité visuelle de la vidéo. Des résultats satisfaisants ont été obtenus avec un coefficient Spearman de 0.97. L'OMQV basé sur les RNAs offre une architecture générique et simple qui peut être utilisée aussi bien dans le contrôle de la qualité d'une vidéo ou d'une image que pour la simple évaluation d'un décodage vidéo numérique.

Dans le chapitre 6, nous utiliserons les résultats de ce chapitre pour démontrer comment un tel outil peut aider à améliorer les performances d'un algorithme de correction d'erreurs dans les images en sortie du décodeur.

4 Evaluation de la Qualité d'une Vidéo par Régression Non Linéaire

4.1 Introduction

Les techniques modernes du décodage vidéo numérique développées cette dernière décennie démontrent des comportements non linéaires qui sont à l'origine de l'apparition de nombreux artefacts visuels sur les images ou les vidéos décodées. Il est plus approprié que la dissimulation de tels artefacts soit effectuée dans une étape post-décodage. Cependant, la plus part des traitements d'images sont implémentés dans un contexte linéaire. Par conséquent, l'architecture des systèmes séquentiels atteint souvent des résultats insatisfaisants lors de la correction d'erreurs de décodage. Dès lors, il serait intéressant d'évaluer le type et le niveau de dégradation de tout artefact visuel durant le processus de décodage.

Dans le chapitre 2, l'OMQV basé sur la classification a permis d'évaluer la qualité d'une image ou d'une vidéo par regroupement en classes, en donnant un niveau de qualité correspondant à l'image ou la vidéo, sur une échelle de 0 à 5. Cette technique a présenté une insuffisance particulière, celle de ne pas fournir un score numérique équivalent au MOS et pouvant le remplacer. Dans le chapitre précédent, l'OMQV basé sur les RNAs a permis d'évaluer la qualité d'une image par estimation d'un score de qualité correspondant au MOS de l'image, sur une échelle de 0 à 5. Cette technique a présenté un avantage quant à la forte corrélation par rapport aux MOS. Toutefois, l'étude d'outils d'analyse statistique tels que la régression non linéaire (RNL) peut aider à essayer d'approcher encore mieux les MOS à travers la modélisation de fonctions mathématiques.

Ce chapitre est organisé comme suit: le paragraphe 4.2 introduit l'outil d'analyse statistique mise en œuvre pour la mesure de la qualité d'une vidéo ou d'une image. Le paragraphe 4.3 décrit des travaux réalisés en [38] ; le paragraphe 4.4 décrit la simulation de l'OMQV basé sur la régression non linéaire et en présente les résultats. Enfin, le paragraphe 4.5 conclut le chapitre.

4.2 Etat de l'art: Utilisation de la régression dans le test et le contrôle

L'utilisation de la régression non linéaire dans l'évaluation de la qualité d'une vidéo ou d'une image est une nouveauté. Cette idée s'inspire de travaux réalisés par Khereddine sur l'utilisation des techniques de régression pour le test et le contrôle des performances des composantes Radiofréquences (RF) [35]. Dans cet article, un modèle autorégressif avec des variables exogènes et un processus d'identification récursive des paramètres dit "boîte noire" sont utilisés dans le but de faire le test et contrôler les performances et le fonctionnement des systèmes AMS (analogue and mixed signal) et RF. La mise à jour continue du modèle est assurée par des algorithmes récursifs d'estimation paramétrique, qui ont l'avantage de permettre une gestion parcimonieuse de la capacité de stockage et de la puissance des ressources nécessaires.

Le test d'un système se base soit sur des mesures (métriques), soit sur des performances du circuit CUT (Circuit Under Test) à tester. Si on considère par exemple un filtre, les performances pouvant être vérifiées sont la bande passante, la fréquence de coupure, etc. Des capteurs sont alors intégrés dans les CUT dans des endroits précis pour récupérer les métriques ou mesures performances utiles au test.

Les techniques de test dit alternatif utilisent un concept de mesures alternatives [57] des performances analogue à celle illustrée dans la Figure 1.1. Dans cette approche, les équations de régression déterminées par identification de modèle permettent d'établir une correspondance entre les variations dans l'espace des métriques de test et les variations dans l'espace des performances de test. Trois enjeux sont alors à considérer :

– Le choix judicieux des mesures : il doit se faire de façon à garantir une bonne corrélation avec les performances à tester;

– Le choix des modèles de régression non linéaire, qui doit se faire dans le but de couvrir l'espace des performances ;

– La définition d'un critère d'optimisation.

La Figure 4.1 illustre le concept des mesures alternatives utilisant le RNL pour le test. Les mesures et les performances obtenues vont alors directement dépendre des paramètres de conception du CUT. La qualité du modèle de régression permettra de savoir si le jeu de test est optimal ou non. Si la qualité du modèle de régression n'est pas satisfaisante, un autre jeu de test est généré et ainsi de suite. Nous remarquons ici que la décision sur le test est prise directement à partir des informations apportées par les mesures.

Ce même travail de recherche présente des méthodes et techniques de régression utilisées pour le contrôle des circuits en fonctionnant 'on' (en temps réel) ou hors fonctionnement en 'off'. Nous nous en sommes également inspirés pour l'élaboration de la boucle de contrôle d'un décodeur (cf. ch. 6).

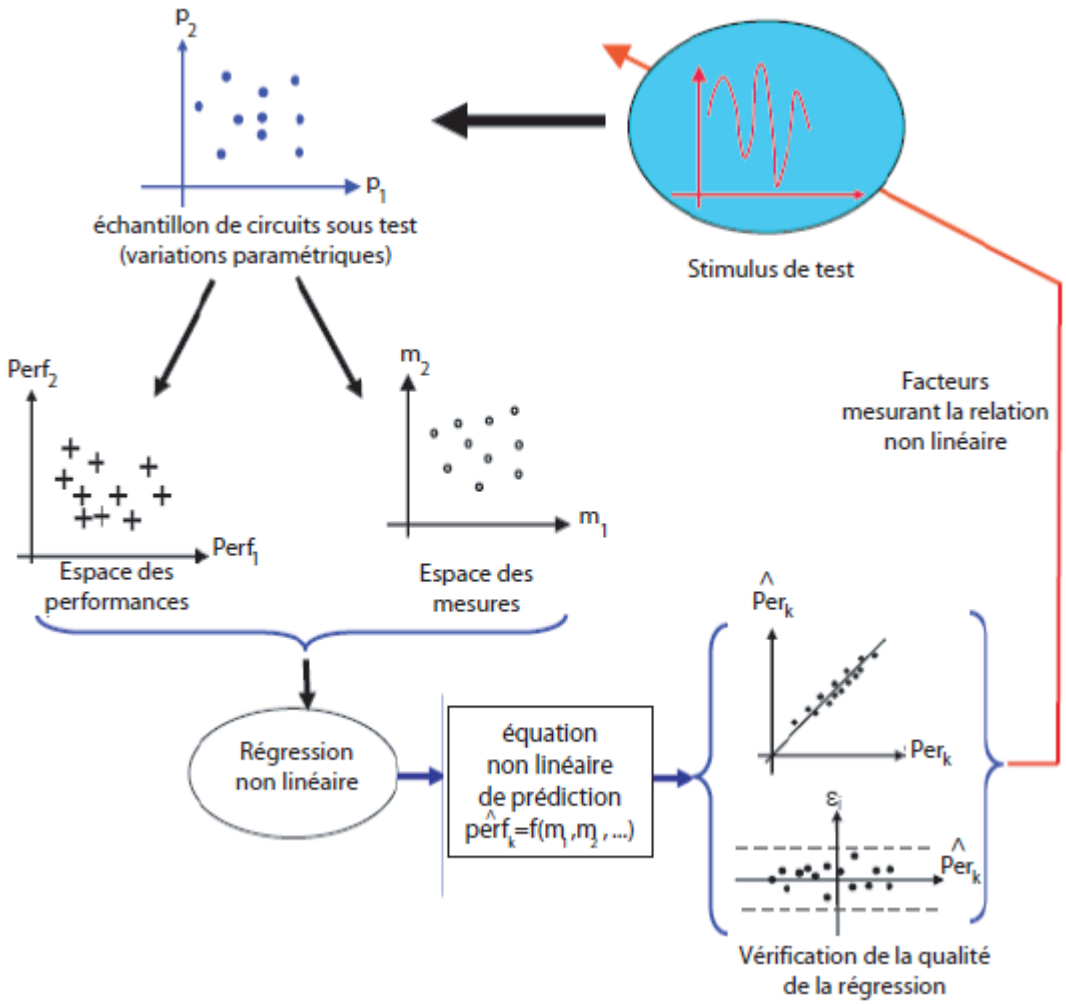


Figure 4.1 Synoptique général de génération de test alternatif [57]. Illustration de la mesure alternative dans le test de circuits. Par similarité avec la Figure 1.1, les mesures représentent les métriques et les performances représentent les scores de qualité.

Ce travail de recherche a permis de développer une méthode efficace pour le test des circuits RF grâce à la régression non linéaire. C'est cette idée qui va inspirer notre technique de mesure la qualité de médias numériques, basée sur la régression non linéaire. Les scores de qualité représentent la fonction coût qui est à prédire à partir des métriques extraites des images.

4.3 OMQV basé sur la Régression

Le reste du chapitre présente une technique de prédiction du score de qualité d'une image ou d'une vidéo basée sur la régression non-linéaire (RNL). Le choix de la régression non linéaire parmi les solutions proposées dans l'état de l'art, est motivé dans la prochaine section.

4.3.1 Généralités de l'OMQV basée sur la régression non linéaire

A. Motivations

L'utilisation de la RNL est motivée par l'envie d'atteindre une meilleure corrélation entre les scores de qualité estimés et le jugement humain exprimé au travers des MOS. La RNL pour avoir été utilisée dans de nombreuses applications similaires [38] offre en effet des possibilités d'approcher au mieux des scores plus proches des MOS que ceux donnés par les RNA dans le chapitre précédents. L'algorithme proposé dans ce chapitre permet en outre d'atteindre des niveaux de corrélation plus élevés avec les MOS, à des moindres coûts quant à la complexité d'implémentation matérielle par rapport à la solution basée sur les RNAs. La non linéarité du model recherché est due à la nature non linéaire de la dépendance entre le MOS et les métriques de qualité (voir figure 2.2 et figure 2.5).

B. Principe

Le premier défi à résoudre quant à la modélisation du problème de régression non linéaire est celui de trouver la fonction (le modèle) qui modélisera le mieux le comportement de la dépendance de la qualité de la vidéo par rapport à chacune des variables dont elle dépend (métriques).

Dans la démarche de recherche des modèles de régression on procède de la façon suivante :

- On crée un premier modèle par déduction de la forme de la courbe de représentation des MOS en fonction des métriques. En remarquant qu'elle est proche de la courbe d'une fonction basique (une hyperbole par exemple)
- On affine le premier modèle créé par transformation (translation, symétrie, …), ce qui nous donne un deuxième modèle.
- On prend un modèle polynomial à partir de l'approximation de la représentation du MOS en fonction des métriques. Ce qui constitue le 3ème modèle.
- Parmi les 3 modèle précédents, on choisi le modèle ayant une plus forte corrélation avec le MOS

Modèle RNL pour la métrique PLR:

Pour le modèle du MOS en fonction du PLR (Figure 4.1), on remarque que la forme de la fonction est proche d'une hyperbole (la fonction inverse par exemple). En modifiant à une constante près son équation mathématique, on obtient un modèle f. Pour obtenir une meilleure précision de l'indice de qualité estimée, nous avons utilisé l'outil Matlab appelé polynomial fitting. Il permet à partir d'un ensemble de points donnés de la courbe d'une fonction, de retrouver un polynôme d'interpolation qui correspond le mieux à cette fonction. La courbe du polynôme d'interpolation est représentée sur la même figure que le modèle initialement estimé afin de pouvoir comparer les deux modèles et d'en sélectionner le meilleur.

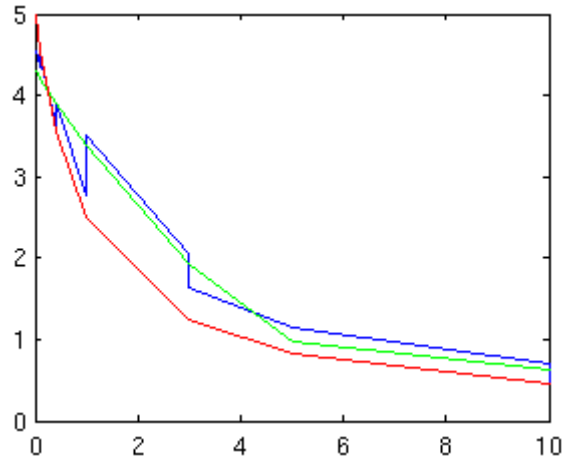


Figure 4.2 Influence du PLR sur le MOS (en bleu) et les modèles de régression associés.

La Figure 4.2 présente l'influence du PLR sur le MOS (courbe en bleu), le modèle initialement créé pour représenter cette dépendance (en rouge) et le polynôme obtenu à partir de l'approximation Matlab (en vert). La figure montre que la forme la plus proche de la courbe représentant le MOS en fonction du PLR est la courbe en vert. Ce point de vue est confirmé par les coefficients de Spearman obtenus en considérant séparément les 2 modèles: SCC = 0,9328 pour le modèle polynomial estimé par Matlab (modélisé en (4.2) par la fonction g) et SCC = 0,928 pour le modèle initialement proposé (modélisé par la fonction f). Il ressort de cette comparaison que le modèle polynomial g est celui qui représente mieux la corrélation entre le PLR et la qualité d'une image.

La fonction f correspond au modèle initialement créé, d'équation (4.1). Elle est une translation de la fonction hyperbolique inverse.

$$f(x) = \frac{1}{x+1} \tag{4.1}$$

$$g(x) = -0.0116x3 + 0.2196x^2 - 1.3872x+4.3776 \tag{4.2}$$

- Modèle RNL pour la métrique PSNR:

La Figure 4.3 présente l'influence du PSNR sur le MOS (en bleu). Le premier modèle (en noir), deuxième modèle (en rouge) et le troisième modèle (en vert) associé à cette dépendance est estimé par la fonction f1, f2 et f3 respectivement (d'équations (4.3), (4.4) et (4.5)):

$$f_1(x) = \left(\frac{x}{18}\right)^2 \tag{4.3}$$

$$f_2(x) = \left(\frac{x-5}{25}\right)^8 + 0.6 \tag{4.4}$$

$$f3(x) = -5.1728.10\text{-}4x3 - 0.0418x^2 + 1.296x - 12.7682 \quad (4.5)$$

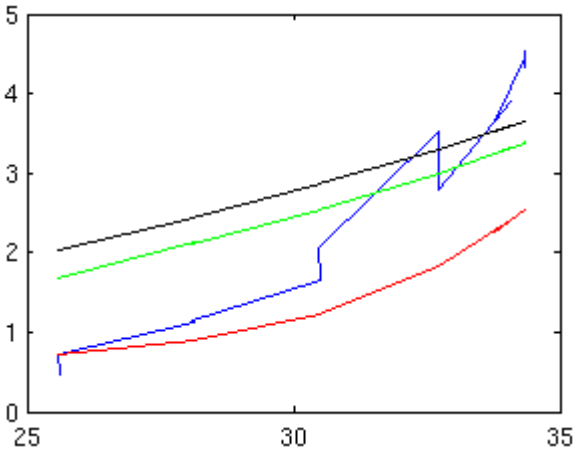


Figure 4.3 Dépendance du MOS sur le PSNR (en bleu) et les trois fonctions f1, f2 et f3 d'estimation respectivement en noir, rouge et vert.

Comme dans l'estimation du modèle du PLR, un premier modèle est proposé et représenté en figure 4.3 (courbe en vert) puis un 2ème (courbe en noir) et le 3ème modèle (en rouge) est déterminé par estimation des fonction polynomiale de la régression entre le MOS et le PSNR.

Les coefficients de détermination de ces trois modèles, R1 = 0,9326, R2 = 0,9326 et R3 = 0,9328 ont été comparés et le premier modèle a finalement été choisi car il correspond à une meilleure corrélation avec les MOS.

- Modèle de RNL de la métrique SI :

Ce modèle a été représenté avec la fonction identité Id à cause du fait que la métrique SI est quasi invariante vis-à-vis des MOS données par les humains.

- Modèle de RNL de la métrique blurMetric:

La Figure 4.4 présente l'influence de la métrique blurMetric sur le MOS (en bleu). Une méthode de modélisation qui implémente un algorithme de calcul du score par régression non linéaire sera mise en place pour cette métrique vu sa forme complexe. L'algorithme sera utilisé pour toutes les métriques et permettra finalement de calculer le score estimé par RNL à partir des 4 valeurs des métriques de la vidéo.

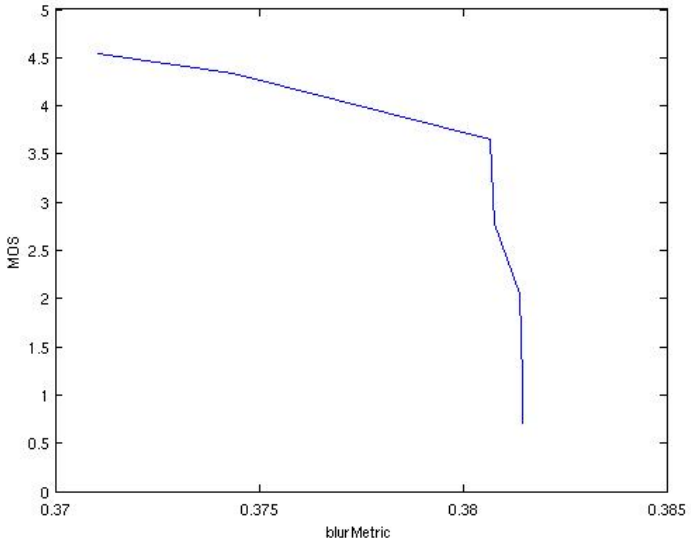


Figure 4.4 Dépendance du MOS (en ordonnées) en fonction de la métrique de flou blurMetric (en abscisse) pour les 6 vidéos sans perte de paquets.

4.3.2 Conception de l'OMQV basée sur la RNL

L'apport majeur de l'OMQV utilisant la RNL par rapport à celui basé sur les RNAs est d'une part le côté pratique de pouvoir jouer sur les modèles et les conditions d'arrêt de l'algorithme pour atteindre une corrélation satisfaisante entre les scores estimés par le modèle de régression et les MOS donnés par les humains ; d'autres parts le contrôle possible sur les influences croisées entre les métriques. La régression non linéaire a été utilisée dans des domaines variés du monde scientifique et a toujours apporté des résultats satisfaisants quant à son application. Elle est surtout utilisée pour l'estimation de fonctions ou des solutions qui modélisent une dépendance entre deux variables.

Dans cette application, il s'agit d'approximer la fonction donnant le score de qualité d'une image ou d'une vidéo en fonction des métriques de qualité extraites de cette d'une image ou vidéo. La base de données utilisée reste la même que dans les 2 chapitres précédents (n=78 séquences vidéos, m=4 paramètres (attributs) de qualité pour chacune des vidéos et M = $(v)_{\mathrm{i,j}}$ la matrice correspondant à ce problème. La colonne j_0 de la matrice M représente le j_0 ème métrique ; et i représente l'index de la i^{ème}séquence vidéo (la i^{ème}ligne) de la matrice.

A chaque itération, l'algorithme considère la métrique ayant le plus grand coefficient de corrélation avec le MOS. Le calcul du maximum Cormax des coefficients de corrélation est donné par (4.6). Cela permet de classer les attributs par ordre de contribution croissante dans l'évaluation quantitative de la qualité d'une vidéo. On représente par $V_{j\ (1\leq j\leq m)}$ un vecteur d'attributs, c-à-d une colonne de la matrice M : V_1 pour la métrique PLR, V_2 pour la métrique PSNR, V_3 pour la métrique SI et V_4 pour la métrique blurMetric.

$$Cor_{max} = \underset{(1\leq j\leq m)}{\mathrm{Max}} \left|Cor(V_j, MOS)\right| \tag{4.6}$$

La détermination du polynôme des moindres carrés est donnée par l'équation (4.7). Notons par j_{max} l'index du paramètre le plus corrélé avec le MOS, le score de qualité approximatif est alors donné par l'équation (4.8). Le vecteur $V_{jmax} = \mathrm{M}(:, \mathrm{j}_{\mathrm{max}})$ correspond à la métrique la

plus corrélé avec le MOS. L'erreur résiduelle Res qui est la différence entre le MOS et le MOS estimé par RNL S est donnée par l'équation (4.9).

$$P = (V_{jmax}T.\ V_{jmax})\text{-}1.\ V_{jmax}T.MOS \quad (4.7)$$

où P est le polynôme donnant les Scores estimés S par les moindres carrés récursifs (MCR).

$$\widehat{MOS} = P.M \quad (4.8)$$

$$Res = MOS - \widehat{MOS} \quad (4.9)$$

L'algorithme se termine si l'une des deux conditions suivantes est vérifiée:

- Si degree(P) $>$ 4
- Si MSE $\leq$ 0.05

A chaque nouvelle itération :

- les nouveaux attributs sont réévalués sous la forme V_{jmax}^2, $V_{jmax}.V_{j\ (j\neq jmax)}$, $V_{j\ (j\neq jmax)}^2 \ldots$ etc.
- Res est considéré comme un vecteur de n éléments, le MSE est donné par l'équation (4.10).

$$MSE = \sum_{i=1}^{i=n} \frac{(\overline{Res} - Res(i))^2}{n} \quad (4.10)$$

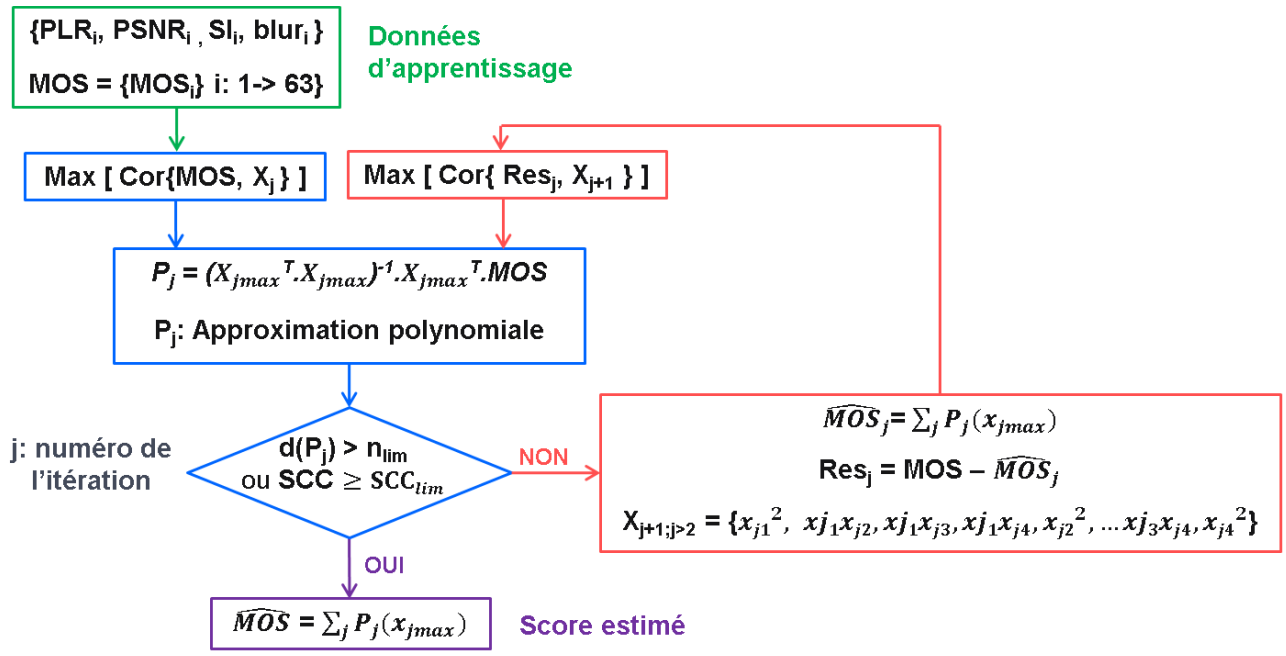


Figure 4.5 Algorithme de prédiction du MOS estimé par régression non-linéaire.

4.4 Modélisation de l'OMQV basé sur la RNL

4.4.1 Algorithme de régression

- Implémentation de la 1ière itération sur Matlab:

a) Calcul des coefficients de corrélation et rangement dans l'ordre croissant:

Les métriques (attributs) initiaux sont stockées dans train1 et les MOS dans target1

```
pour i=1:2
   cor_tt(i)=abs(corr(target1,train1(:,i)));
fin
[cor_sort,i_sort] = sort(cor_tt);
```

b) Récupération des paramètres significatifs (si le coefficient est supérieur à 0,5) et de leur index :

```
number_imp=sum(cor_sort>0.5);
M_index = 1:length(target1);
```

c) Initialisation :

```
target1_store=target1;
u_chap_iteration1=zeros(length(target1),1);
```

d) Boucle d'estimation par régression :

```
SCT=sum((target1-mean(target1)).^2);
pour i=1:number_imp
   i_max=i_sort(end+1-i);    (l'indice du paramètre le plus corrélé au MOS)
   p(i,:) = polyfit(train1(:,i_max),target1_store,3);  (le polynôme des MCR)
   u_chap = polyval(p(i,:),train1(:,i_max));  (Les scores estimés par RNL)
   residu = target1_store - u_chap;   (Le Résidu)

   SCE=sum((residu).^2);
   r2_1(i)=1-((SCE) / (SCT));
   target1_store=residu;
   u_chap_iteration1=u_chap_iteration1+u_chap;
   u_chap_arry(i,:)=u_chap_iteration1;
fin
```

- Implémentation de la nème iteration sur Matlab:

a) Calcul des coefficients de corrélation et rangement dans l'ordre croissant:

```
  r2_n=1-((SCE) / (SCT))
  c=1;
pour i=1:1
   pour j=i+1:2
      varn-1(:,c) = train1(:,i).*train1(:,j);
      cor_rv(c)= abs(corr(residu,varn-1(:,c)));
      c=c+1;
```

```
    fin
fin
[cor_rv_sort,i_rv_sort] = sort(cor_rv);

b) récupération des paramètres significatifs (si le coefficient est supérieur à 0,1) :
number_imp=sum(cor_rv_sort>0.1);

c) Initialisation
u_chap_iterationn=u_chap_iteration1;


d) Boucle d'estimation par régression
pour i=1:number_imp
   in_max=i_rv_sort(end+1-i);    (l'indice du paramètre le plus corrélé au MOS à
l'itération n)
   p(in-1_max+i,:) = polyfit(varn-1(:,in_max),target1_store,3);  (le polynôme des MCR)
   u_chap = polyval(p(i_max+i,:),var1(:,in_max));  (Les scores estimés par RNL)
   residu = target1_store - u_chap;

   SCE=sum((residu).^2);
   r2_n(i)=1-((SCE) / (SCT));
   target1_store=residu;

   u_chap_iterationn=u_chap_iterationn + u_chap;
   u_chapn_arry(i,:)=u_chap_iterationn;
fin
```

- Résultats de la 1ière itération:

Le polynôme d'interpolation P est du 3ème degré. Le coefficient de détermination obtenu après cette première itération est SCC_1= 0.929; l'erreur quadratique moyenne est MSE_1= 0.127. La Figure 4.6 présente la comparaison entre les MOS et les scores estimés par l'OMQV basé sur la RNL après la première itération.

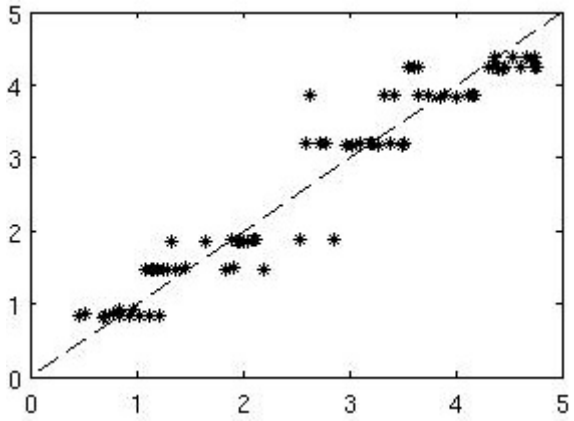


Figure 4.6 MOS (en abscisse) vs Scores estimés par RNL (en ordonnées), 1ère itération.

- Résultats de la 2ième itération:

Le polynôme d'interpolation P est du 4ème degré. Le coefficient de détermination obtenu après cette deuxième itération est SCC_2= 0.939; l'erreur quadratique moyenne est MSE_2= 0.109. La Figure 4.7 présente la comparaison entre les MOS et les scores estimés par l'OMQV basé sur la RNL après la deuxième itération.

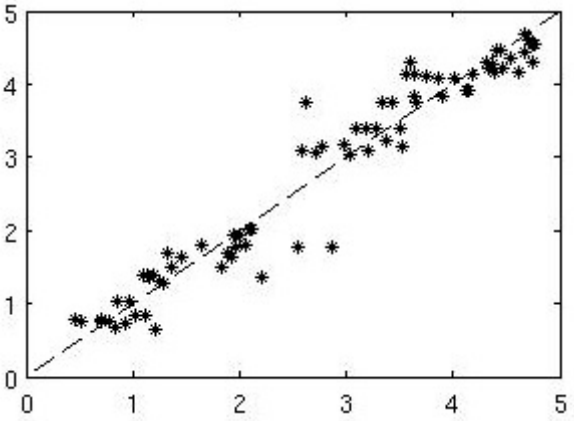


Figure 4.7 MOS (en abscisse) vs Scores estimés par RNL (en ordonnées), 2ème itération.

- 3ième itération:

La Figure 4.8 présente les résultats de la troisième itération. L'algorithme prend fin à cette itération, car le degré du polynôme à cette itération est de 5. On retrouve donc les performances de l'OMQV basé sur la RNL : Le coefficient de détermination SCC_3= 0.946 et l'erreur quadratique moyenne MSE = 0.096.

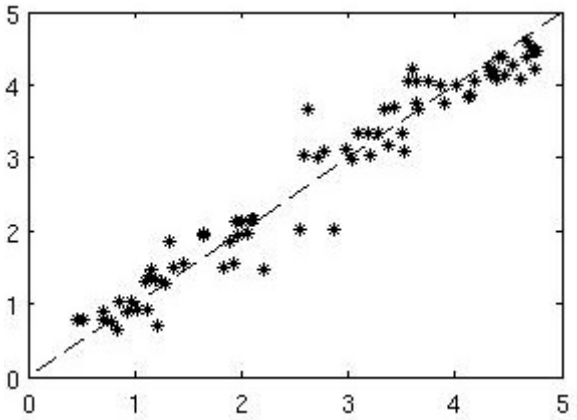


Figure 4.8 MOS (en abscisse) vs Scores estimés par RNL (en ordonnées), 3ème itération.

4.4.2 Résultats et Discussions

Les résultats obtenus démontrent que l'OMQV base sur la RNL évalue les séquences vidéo d'une façon très proche du jugement humain. 80% des vidéos de la base de données ont été soumises à une évaluation par l'OMQV basé sur la RNL et les scores estimés ont été comparés aux MOS. La corrélation satisfaisante obtenue à partir de cette comparaison avec un coefficient de SCC = 0.946 avec les MOS prouve que cette méthode utilisant le RNL obtient des performances sur l'EQV très proches du système visuel humain (SVH).

On peut constater que dans la Figure 4.2 et la Figure 4.3, les scores donnés par les humains (MOS) sont plutôt localement discontinus entre les niveaux de qualité 2 et 3, par rapport aux autres niveaux de qualité de vidéo. Cette discontinuité se traduit aussi dans la Figure 4.6, Figure 4.7 et Figure 4.8 où les valeurs des scores estimés dans l'intervalle [2 ; 3] des niveaux de qualité restent localement situés aux extrémités. Ceci peut être interprété par le fait que le jugement humain est très dispersé dans ce niveau de qualité intermédiaire ou plutôt que lors de l'évaluation, les observateurs ont tendance à diviser l'échelle des niveaux de qualité en deux intervalles, l'intervalle [0 ; 2] des vidéos de qualité "insatisfaisante" et l'intervalle [3 ; 5] des vidéos de qualité satisfaisante. Ceci correspond plutôt à une sorte de classification.

Une trame de la séquence vidéo "mobile" est représentée dans la Figure 4.9 avec trois différents taux de perte de paquets. Elle a été évaluée par l'OMQV basé sur la RNL. Le SI de cette séquence vidéo est de 11 et le TI de 10. Les autres métriques des 3 trames correspondantes à ces 3 niveaux de PLR ainsi que les valeurs du MOS estimé par RNL et les MOS sont données dans le tableau 4.1.

Tableau 4.1 Métriques, MOS estimés et MOS pour les frames en Figure 4.9.

PLR	PSNR	blurMetric	MOS	MOS estimés par RNL
PLR = 0	28.29	0.219	4.71	4.626
PLR = 0.4%	27.8	0.219	4	4.01
PLR = 5 %	23.16	0.219	1.46	1.72

Figure 4.9 Vidéo "mobile" :de gche à dte: réf. ; avec un PLR = 0.4% ; et avec un PLR=5%.

Importance des métriques TI et SI sur l'évaluation des vidéos:

Comparons les performances obtenues sur l'évaluation des vidéos par celles qu'on obtiendrait en ne considérant pas la métrique SI. La Figure 4.10 montre la comparaison entre les MOS et les scores obtenus par RNL après simulation en ne considérant pas SI, c'est à dire en considérant uniquement le PLR et le PNSR. Les performances obtenues sont les suivantes :

- $SCC_{sans\ SI}$ = 0.929< R^2 et $MSE_{sans\ SI}$= 0.127> MSE

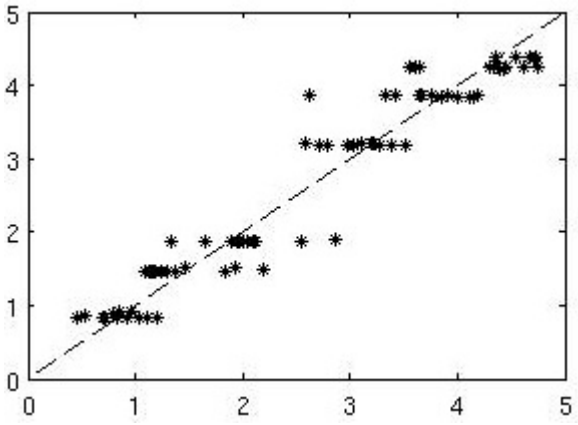


Figure 4.10 Scores estimés (en ordonnées) et MOS (en abscisse) en l'absence des attributs SI et TI (pour comparaison avec la Figure 4.8).

Cette comparaison montre clairement que de meilleures performances sont obtenues en considérant les attributs SI et TI. Ce résultat est logique puisque des métriques additionnelles apportent plus d'informations sur la perception visuelle et les effets de distorsion sur le contenu de deux séquences vidéo distinctes. On aurait pu considérer d'autres métriques dans ce projet afin d'obtenir de meilleurs résultats mais les 4 métriques PLR, PSNR, SI et blurMetric ont été les seules métriques retenus parce qu'elles sont faciles à extraire sur la base de données considérée (en particulier pour PLR, SI et TI) et parce qu'elles (en particulier les métriques PSNR et blurMetric) mesurent les niveaux de détérioration des 2 artefacts (bruit et flou) les plus connues dans le décodage MPEG qui fait l'objet de notre étude.

Combinaison des techniques RNL et Classification k-NN :

La combinaison de la RNL et de la classification (cf ch. 2) par classes de qualité (cf. Table 6.1) de 62 vidéos de la base de données considérée a donné les résultats de la Figure 4.10.

Ces résultats montrent clairement que l'évaluation visuelle –et faite selon le jugement humain - ne considère pas de niveau intermédiaire (niveau de qualité passable) dans l'échelle de qualité des vidéos. Il distingue les 64 vidéos en deux groupes : Les vidéos de qualité mauvaise ou pire qui peuvent être classé dans un même groupe qu'on appellera "qualité insatisfaisante"; et les vidéos de qualité bonne ou excellente, qui peuvent également être classées dans un même groupe qu'on dénommera "qualité satisfaisante".

Des travaux futurs pourront étudier et considérer d'autres métriques telles que la texture, les effets de bloc ou le flou afin d'augmenter les performances de l'OMQV et sa corrélation avec le jugement humain.

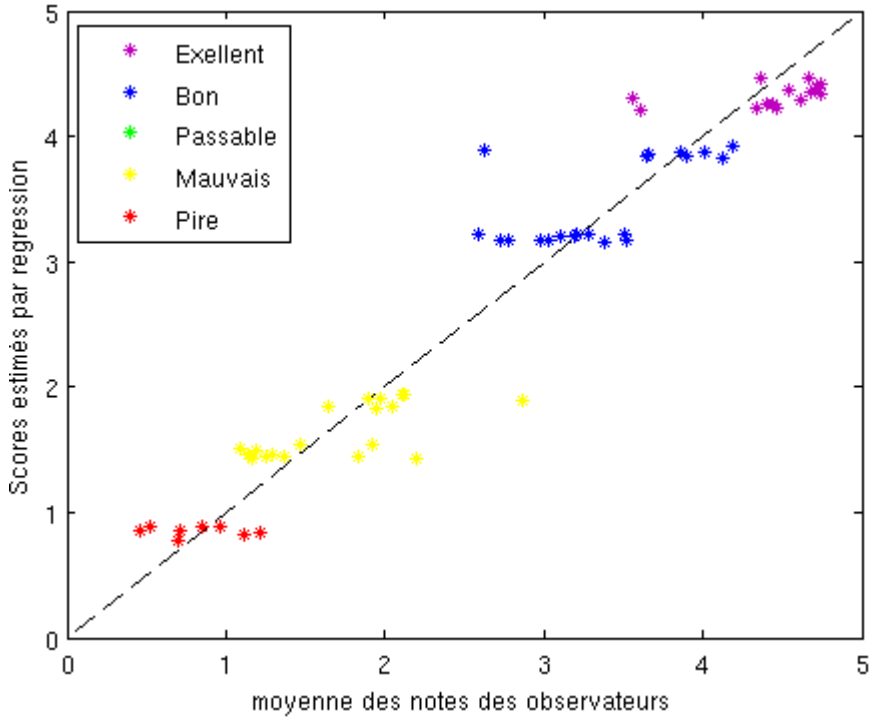


Figure 4.11 Classification et mesure par régression des vidéos de la base de données

L'implémentation bas niveau de l'OMQV sur FPGA pourrait faire l'objet de prochains travaux de recherche à la suite de cette thèse. Le code Matlab embarqué a déjà été édité notamment pour l'OMQV basé sur les RNAs, montrant un niveau de complexité élevé, raison pour laquelle la solution RNL a été préférée. La solution RNL quant à elle serait d'autant complexe que le degré du polynôme d'interpolation augmente, car ça augmente le nombre de transistors. C'est pourquoi nous nous sommes limités à un niveau de complexité correspondant à un maximum de 4 pour le degré du polynôme des moindres carrés utilisé dans la RNL. La condition sur la valeur minimale de MSE est aussi un élément qui joue sur la

complexité matérielle du système. En effet plus l'erreur résiduelle minimale est basse, plus le nombre d'itération est élevé, et donc le système devient plus complexe.

4.5 Conclusion

Dans ce chapitre un outil de mesure de la qualité d'une vidéo (OMQV) basé sur la régression non linéaire a été élaboré. Cet outil évalue la qualité d'une vidéo en lui attribuant un score numérique sur une échelle de 0 à 5, correspondant au niveau de qualité de cette vidéo. La valeur du coefficient de corrélation (SCC = 0.9406) avec les MOS (moyennes des opinions des sujets (humains)) démontre des résultats très satisfaisants quant à la cohérence de la méthode avec le jugement humain. Une autre application de la régression non linéaire est illustrée plus loin (chapitre 6) pour la priorisation entre deux ou plusieurs types d'artefacts affectant une vidéo.

Des travaux futurs pourraient définir des métriques complémentaires (flou, texture, …) à ajouter aux métriques considérées dans ce projet, afin d'augmenter la précision sur la mesure de la qualité des vidéos et d'accroitre la corrélation avec les MOS.

5 Méthodes de diagnostic et la correction d'artefacts visuels

5.1 Introduction

Dans les premiers chapitres nous avons développé des techniques pour l'évaluation de la qualité vidéo en utilisant la méthode avancée d'analyse statistique et l'intelligence artificielle. Dans le reste du manuscrit, nous utiliserons les résultats de ces travaux pour donner une contribution à l'amélioration de la qualité visuelle des vidéos décodées par les décodeurs numériques de la famille MPEG. En effet, une fois que le niveau de qualité d'une vidéo donnée connue, il est nécessaire de détecter les sources de distorsions qui affectent cette vidéo afin de corriger les erreurs et de réaliser une amélioration visuelle de la qualité de cette vidéo.

La dissimulation d'artefacts visuels étant une démarche multidirectionnelle, il importe de diagnostiquer au préalable les principaux artefacts visuels à la sortie des vidéos. Toutefois, une certaine interdépendance entre ces artefacts fait apparaitre des effets de masquage. Ces effets de masquage rendent encore plus complexe les tâches de correction d'erreurs.

Dans l'état de l'art, bon nombre de techniques ont utilisées l'implémentation d'algorithmes de correction d'erreurs basée sur des informations de type spatiaux ou temporelle pour la correction des erreurs dans une image. La dissimulation d'artefacts peut aussi se faire sur la base des types de source ayant causé ces artefacts. Ainsi, la compression jpeg peut causer des artefacts de type effets de bloc, ou bien du type bruit. De même, les artefacts de type flou peuvent en général être causé par un déplacement de l'objectif (ou de la lentille) ou bien par un mauvais ajustement du zoom de la lentille de l'objet de capture (camera par exemple). Dès lors, le diagnostic d'erreurs s'avère être lié aux techniques de correction de ces mêmes erreurs. Certains algorithmes correcteurs vont ainsi développer des actions visant à réduire les artefacts visuelles qui apparaissent à la sortie des images par des opérations de convolution avec des filtres adaptatifs [62] selon le type d'artefact alors que d'autres vont plutôt agir sur l'information spatiotemporel ([5] et [26]) de l'image ou de frames adjacentes sur une même vidéo. Dans ce chapitre, nous présentons des résultats pertinents précédemment obtenus sur le diagnostique et de correction d'erreurs dans le domaine du traitement d'image.

Le reste du chapitre se présente comme suit : Le paragraphe 5.2 présente les généralités sur les artefacts visuelles. Quelques techniques de dissimulation d'artefacts visuelles sont décrites dans le paragraphe 5.3 et le paragraphe 5.4 conclut le chapitre.

5.2 Généralités sur les artefacts

Au regard des avancées réalisées jusqu'ici dans le domaine du traitement d'images et de la vidéo, beaucoup d'étapes ont été franchies mais la course continue vers la perfection dans la qualité de la vidéo capturée, transmis et décodé. De plus en plus, le challenge dans le domaine du traitement de données multimédias est à l'optimisation des vidéos reçues sur les équipements numériques allant de petits écrans embarqués sur les appareils mobiles à grand écran de télévision. L'idéal pour tout utilisateur regardant une vidéo est d'obtenir une qualité visuelle le mieux possible satisfaisant. Ce qui implique une dissimulation optimale de tous les artefacts perceptibles. Par conséquent, il est nécessaire de détecter, diagnostiquer et dissimuler autant que possible les artefacts visibles sur une vidéo décodée.

La Figure 5.1 présente l'architecture basique d'un décodeur MPEG. Le flux de données entre les appareils de traitement vidéo implique nécessairement le passage à travers un décodeur vidéo numérique. Nous avons tous certainement connu le désagrément de visionner une vidéo de mauvaise qualité lorsque l'on souhaiterait regarder une scène capturée alors qu'on préfèrerait la regarder avec la plus grande qualité visuelle possible. Certaines sources de ces artefacts surviennent en amont de la boucle de codage et peuvent être amplifiées de telle sorte que la distorsion devient donc plus importante à la sortie de la boucle de décodage.

Nous allons maintenant nous intéresser aux aspects liés à la qualité de la vidéo décodée en amont comme en aval de la boucle de décodage. Certaines causes de ces défauts dans la qualité d'une vidéo sont connues du grand public tel que flou. D'autres causes telles que le bruit gaussien sont moins perceptibles à l'œil nu, mais avec autant d'effets néfastes sur la qualité de la vidéo. La suite d'exemples d'artefacts fournie ici n'est pas exhaustive et dans la pratique, d'autres variantes de chaque type d'artefact peuvent se produire. Restreindre le nombre d'artefacts à quatre était nécessaire parce que les expériences qui estiment les distorsions et l'étude faite sur ces artefacts nécessitent une grande quantité de données et un nombre raisonnable d'images originales.

5.2.1 Principaux types d'artefacts

Il existe plusieurs types d'artefacts visuels à la sortie des vidéos. Ces types d'artefacts sont de nature différente. Cette nature se distinguera par l'apparence et l'aspect de l'artefact sur la vidéo déformée. Dans la norme MPEG, l'encodage numérique découpe l'image en petits blocs de forme rectangulaire appelé Macroblocs. Ces Macroblocs ont une taille de 16x16 pixels et sont constitués de 4 blocs de 8x8 pixels. Dans le flux de données des décodeurs de la famille MPEG, les informations sont transmises sur le contenu de ces blocs. Ces effets de bloc sont principalement causés par deux phénomènes : La mauvaise quantification de l'image numérique et la compensation de mouvement des pixels lors de la reconstitution de l'image.

La mauvaise réception du signal numérique induit des erreurs sur les données reçues. Ainsi, le contenu des Macroblocs peut entre-autre être corrompu. On se retrouve alors avec des blocs mal placés sur l'image, avec un contenu qui n'est plus cohérent (voir Figure 5.4).

Le débit du flux MPEG peut quelques fois être limité pour des soucis de gain en bande passante et de budget notamment chez les opérateurs de chaine de TV. Avec cette limitation, les informations concernant le contenu des Macroblocs sont limitées. On se retrouve avec des Macroblocs dont le contenu diffère de l'image originale. Ce problème apparait le plus souvent quand l'image est détaillée avec beaucoup de mouvement. Par exemple la surface de l'eau, ou les spectateurs d'une retransmission sportive. Dans ce cas, le nombre d'information à envoyer est supérieur au débit alloué.

Dans ce paragraphe, nous identifions les principaux artefacts visuels connus dans le décodage MPEG. Ce sont entre autre: les suroscillations, l'effet de bloc, le flou et le bruit.

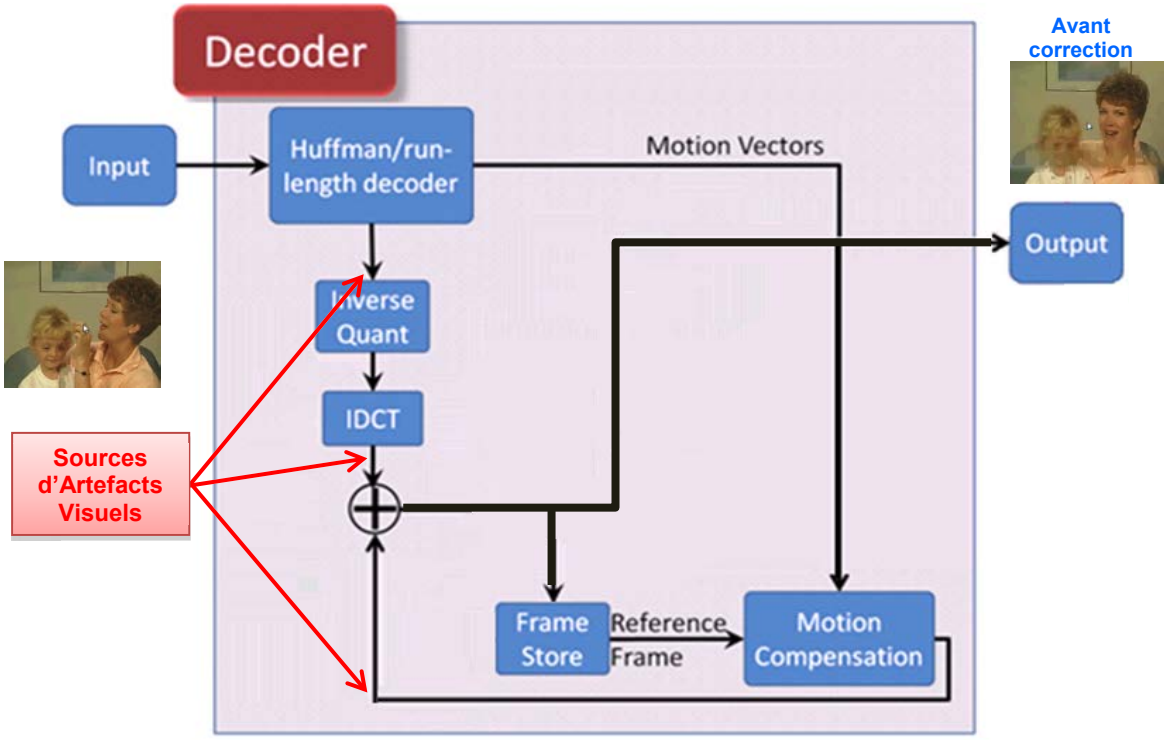


Figure 5.1 Architecture d'un décodeur MPEG: Identification des sources d'artefacts.

A. Les suroscillations

Les suroscillations peuvent apparaître lorsqu'une déconvolution est effectuée dans certaines limites que la bande passante dans le domaine fréquentiel ou des ondulations dans la fonction sinc dans le domaine temporel. Ils apparaissent visuellement comme des bandes spectrales près des bords (Figure 5.2 image de droite). Les suroscillations de l'image restaurée, J3, se produit le long des zones de contraste d'intensité forte à l'image et le long des bords de l'image. Cet exemple montre comment réduire des suroscillations en spécifiant une fonction de pondération.

Le mot "suroscillations" est utilisé à cause de l'effet visuel visible autour des contours. On remarque un écart de taux d'oscillation autour d'une transition brusque à proximité de l'entrée, telle que des oscillations d'une cloche qui vient d'être sonnée. Il existe différentes causes pouvant produire des discontinuités autour des bords dans une image et faire apparaître ces « suroscillations ». Ces discontinuités peuvent se produire sur l'amplitude. L'amplitude est alors impactée au niveau des bords de l'image ou à l'intérieur même de l'image, ou même au niveau de la luminance de l'image.

Figure 5.2 Frame “Foreman”: à gche l’image d’origine obtenue de [17], au centre l’image flou, à dte l’image restaurée à partir de l’image flou, par défloutage à l’aide d’une déconvolution avec un filtre gaussien. On peu remarquer l’effet visuelle de “suroscillations” sur cette dernière frame.

B. L’Effet de bloc

L'un des artefacts vidéo les plus perceptibles et les plus courants de nos jours est l'effet de bloc car la perception visuelle humaine est très sensible aux contours de blocs et aux régions planes de l'image (voir Figure 5.3 b). Cette comparaison montre qu’une compression peu judicieuse peut causer l’effet de bloc dans une image.

Dans le décodage vidéo numérique MPEG, de récents algorithmes de codage vidéo comme compensation de mouvement prédiction, ou les coefficients DCT peuvent produire des effets de bloc quand ils sont appliqués dans le codage à faible débit binaire. En effet, dans un tel mécanisme, la prévision basée sur des blocs ou le traitement de quantification détruisent la correspondance entre blocs voisins. En conséquence, des effets de bloc peuvent apparaître immédiatement aux limites de blocs ou sous-blocs de pixels comme le montre la Figure 5.3.

En outre, les effets de bloc se propagent le long des trames voisines lors du décodage d’une séquence vidéo reconstruite. Les effets de bloc dans les trames précédentes peuvent en effet être propagés à la trame courante dans le processus de reconstruction de la vidéo (voir Figure 5.4).

Figure 5.3 a: Image originale, absence de blocs b: Image jpeg compressée à 10%.

Une trame de la séquence vidéo "mobile" : (a) : l’image originale obtenue de [17], (b) : image décodée par un décodeur en format jpeg avec un taux de compression de 6%.

Figure 5.4 Trame de la séquence vidéo "foreman" : A gauche : l'image originale obtenue de [17]; Au milieu : image de la même séquence vidéo reçue avec un taux de perte de paquets de 0.4%. A droite : image de la même séquence vidéo reçue avec un taux de perte de paquets de 5%.

C. Le flou

Dans le traitement de l'image numérique, le flou est sans doute l'artefact le plus remarquable. Il apparait après un traitement d'image par convolution avec un filtre passe bas. Le flou est causé par une perte de la texture ou de précision lors du décodage d'une image (Figure 5.2 image située au milieu). Le flou se produit lors d'un traitement d'image ou de vidéo en ondelettes à faible débit et est très présente dans la compression JPEG2000. Il peut être causé par la poursuite oculaire d'objets lors d'un déplacement trop rapidement, la mise en forme d'images en vu d'une adaptation à la résolution initiale, la réponse active de pixels d'image en couleur, une perte d'informations lors de la compression etc... Dans la plupart des cas–notamment pour des flous trop accentués – il peut s'agir d'une combinaison de ces différentes causes.

Dans le traitement numérique des vidéos, l'artefact flou peut souvent être causé par la mauvaise quantification des coefficients DCT ou une mauvaise compression lors du décodage (MPEG par exemple). C'est dans ce type de cas que les effets de bloc et flou ont la même origine. Ceci peut être observé dans la Figure 5.3 b, où les effets de bloc et le flou sont causés par un mauvais taux de compression de l'image.

D. Le bruit

Les artefacts de bruit comme leur nom l'indique sont des artefacts causés par la présence d'une grande quantité de bruit (voir Figure 5.5), alors que le bruit sur une image traitée ou un cadre peut être définie comme toutes caractéristiques indésirables. Un peu de bruit sur une image n'est pas nocif et peut dans la plupart des cas, être acceptable pour effectuer une bonne qualité de l'image. Certains Procédé de traitement d'image ajoutent d'ailleurs une grande distorsion sur l'image à traiter, ce qui dégrade un peu la qualité de l'image.

Le bruit a toujours été l'un des principaux problèmes dans le domaine du traitement numérique de l'image. Le bruit peut être interprété comme une répartition hétérogène et inégale des taches de couleur avec une faible fréquence. Les effets nocifs du bruit sont plus dommageables lorsqu'ils sont combinés avec d'autres artefacts d'image en couleur. Le bruit peut prendre l'apparence des taches, des données manquantes, les modes aléatoires ou ordonnés et d'autres variations produisant un regard confus sur une image ou distractions imperfections visuelles. Concernant la compression numérique, existe au moins trois principales catégories de bruit:

Figure 5.5 Une image tirée de [49] avec le côté gauche touchées par des artefacts de bruit et le côté droit exempt de bruits.

- Le bruit emplacement fixe: ce qui correspond à des artefacts toujours situés au même endroit dans l'image, avec des positions prévisibles.

- Le bruit apparaissant au hasard: Le bruit aléatoire produit, des variations de pixels isolés discrets ou «pointes» et donne une image d'apparence "sel et poivre" (Figure 5.5).

- Le bruit cohérent: Correspond à artefacts introduits par des signaux électroniques erronés produits par exemple par le fonctionnement des instruments à bord d'un vaisseau spatial lors de prises ou d'observations d'images.

5.2.2 Détection d'Artefacts

Le domaine scientifique du traitement d'images et de la vidéo a connu de grands progrès dans la détection de la source d'erreurs dans les vidéos décodées ces dernières années. Nous allons étudier quatre travaux de recherche qui ont été réalisés dans ce domaine au cours des dernières décennies.

Dans le papier [64], T. Vlachos a développé une méthode pour la détection des effets de bloc sur les vidéos. La méthode qu'il a développée est pratique pour les applications en temps réel puisqu'elle n'a pas besoin de la vidéo de référence. Dans le papier [24], M. Farias a étudié la détectabilité des effets de bloc et flou et leur nuisance dans une vidéo décodée par le décodeur MPEG-2. Une approche pour la détection du flou basée sur la classification est décrite en [53]. L'évaluation de la qualité de l'image selon une technique dite aveugle décrite dans [46] élabore des études statistiques et des distorsions synthétiques pour détecter et prioriser les sources d'erreurs au sein de l'image.

Détection des effets de bloc dans des vidéos compressées

Considérant que les effets de bloc sont les plus importants dans les artefacts liés à la famille des décodeurs MPEG, Vlachos [64] a élaboré une approche sans référence pour les aider à leur détection dans une vidéo décodée. Sa méthode a l'avantage d'être sans-référence et, par conséquent utile pour les réseaux de communication où le côté récepteur nécessite une grande qualité de la vidéo reçue. En outre, la méthode est d'un grand intérêt pour les applications en

temps réel. Il a appliqué des transformations rapides dans le domaine de fréquence sur des échantillons de blocs 8x8 dans une trame pour chaque image de la vidéo compressée.

Un filtre de Hamming a été appliquée à chaque fenêtre de l'élément de l'échantillon de manière à faire disparaitre les limites du tableau et de conserver l'intégrité des éléments situés au centre. Un échantillon de sous-image situé à un angle de l'échantillon est corrélé aux trois autres sous-images situées aux angles du sous-bloc pour obtenir la mesure de la correspondance entre les limites des blocs voisins.

La détection des effets de bloc consiste alors à évaluer la combinaison des plus hauts sommets des couples de l'échantillon à travers une fenêtre d'image. Le pic est considéré comme une fonction impulsion de Dirac calculée à partir des coordonnées de pixels dans les tableaux et les mesures de corrélation entre les blocs adjacents.

Détectabilité et impact des artefacts de synthèse des effets de bloc et du flou

Cette méthode développe des artefacts de bloc et de flou synthétiques, plus faciles à étudier que les artefacts naturels (distorsions MPEG), mais avec les mêmes effets sur la qualité visuelle des vidéos. Le travail est axé sur les distorsions liées à au décodeur MPEG-2 en supposant certaines propriétés bien vérifiées par les artefacts synthétiques dans les autres afin de rester étroitement lié aux artefacts de compression.

Pour les artefacts synthétiques d'effets de bloc, une fenêtre d'image est d'abord divisé en sous-blocs de 8x8 blocs et la moyenne de chaque bloc est calculée, ainsi que la moyenne des 24x24 blocs ayant le bloc considéré comme le centre, puis la différence entre les deux moyennes est ajouté à chaque bloc de la fenêtre initiale.

L'artefact synthétique de flou est introduit sur une fenêtre d'image par l'application d'un filtre passe-bas de type FRI (Filtre à Réponse Impulsionnelle) à deux dimensions. L'intensité du flou peut être régulée en changeant la fréquence de coupure des différents filtres.

Une superposition des deux types d'artefacts (l'effet de bloc et le flou) nommée artefact blocky-blurry (bloc-flou) a ensuite été obtenue en combinant les coefficients des effets de blocs et ceux des effets de flou. Par ce moyen, les résultats obtenus sont bien proches des artefacts qu'on retrouve dans le décodeur MPEG-2. Un seuil est défini à partir duquel une valeur de l'artefact bloc-flou pourra être considéré comme "Très mauvais" et la comparaison des résultats obtenus comparés avec l'évaluation subjective a montré une très bonne corrélation.

Détection des artefacts de flou dans la compression d'images JPEG2000 à l'aide de classification

Dans ce travail de recherche, l'auteur a créé une méthode de détection des artefacts de flou basée sur la classification. La technique est inspirée de son travail précédent référencé dans [53] et consacré à la détection de textures et d'artefacts utilisant des images codées en ondelettes. En assumant le fait que la localisation des régions floues se fait autour de textures bien conservés principalement pour des artefacts liées à la compression Jpeg2K, il a d'abord détecté les régions texturées en calculant une différence de gaussienne dénotée DOG (Difference of Gaussian) et une différence de compensation gaussien DOOG (Difference of Offset Gaussian) sur les filtres. Dans une deuxième étape, il utilise un algorithme K-mean pour appliquer une segmentation dans les régions texturées à travers l'image. Chaque zone texturée a ainsi été segmentée en des segments adjacents. L'algorithme de Classification a consisté à considérer un seuil S de différence entre la moyenne du segment de source et la moyenne du

segment cible et classifie comme artefacts les valeurs au-dessus et comme texture bien préservé celles en-dessous.

Mesure aveugle des artefacts de bloc en images

Dans [46], une évaluation de la véracité et de l'intégrité des artefacts basée sur l'identification de l'image. L'expérience a consisté à extraire des données statistiques à partir des coefficients de sous-bande d'image. Ces coefficients de sous-bande sont calculés à partir des réponses en passe-bande orientées obtenues par application d'une transformée en ondelettes des images déformées. Le modèle de mélangeur à l'échelle gaussienne a été utilisé pour calculer les coefficients en ondelettes.

Cinq distorsions ont été prises en compte. Ce sont la compression jp2k, la compression JPEG, le bruit blanc, le flou gaussien et l'évanouissement rapide. Les vecteurs sont alors élaborés à partir des données statistiques extraites des images déformées. L'expérience était composée de deux parties: Dans une première partie de l'expérience, la probabilité à laquelle l'un des cinq modes de distorsions impact de la dégradation de l'image a été calculée. Dans la seconde partie, les vecteurs caractéristiques ont été identifiés à un indice de niveau de qualité pour chacun des cinq types de distorsion.

Ce travail est très proche de notre approche pour la détection de la source d'erreurs dans les vidéos décodées avec la norme MPEG. En fait, l'utilisation des fonctions statistiques pour déterminer le type de distorsion le plus nocif sur la vidéo à traiter est le principal point commun avec notre projet. En outre, les cinq catégories de distorsion étudiées ici vont également nous intéresser, car ils représentent les principales distorsions liées à la famille de décodeurs MPEG.

Une fois les principaux artefacts du décodage h.264 définies, nous allons à présent présenter des méthodes de dissimulation de ces artefacts, connues plus généralement sous le nom d'algorithmes de correction d'erreurs.

5.3 Quelques Techniques de dissimulation d'Artefacts Visuelles

Il existe 2 principales approches pour la correction d'erreurs de décodage. Une première se base sur les informations spatio-temporelles contenues dans les trames adjacentes à la trame à corriger au sein d'une même séquence d'images. La seconde se base sur le type d'artefact à corriger. Cette dernière adapte l'algorithme de correction d'erreurs au type de comportement de l'erreur alors que la première famille remplace en générale un sous-bloc erroné dans une trame (image) par un autre sous-bloc se trouvant dans une trame ou une tranche (de macroblocs) voisine.

Ce paragraphe, présente quatre travaux réalisés sur la réduction d'artefacts : l'élimination Spatio-temporelle d'Artefacts ; Un algorithme avancé de dissimulation d'Artefacts pour des intra-trames ; Un algorithme de dissimulation d'erreur basé sur le décodage et l'élimination des effets de bloc dans les Systèmes de Décodage Vidéo basés sur les blocs.

5.3.1 Elimination Spatio-temporelle d'Artefacts Vidéo avec un mode de sélection Perceptivement Optimisé

Ce travail de recherche crée une méthode (voir [5]) pour l'amélioration de la dissimulation d'artefacts visuels. La technique proposée emploie estimation du vecteur de mouvement, l'interpolation de contours préservés et l'analyse / la synthèse de la texture. La dissimulation temporelle peut présenter de meilleurs résultats avec une bonne précision sur le vecteur de mouvement, tandis que la dissimulation spatiale si elle fournit habituellement des estimations imprécis des Macroblocs manquants (ou erronés) est donc préférable à une estimation temporelle inexactes. La particularité de cette méthode est de combiner, dans certains cas, à la fois interpolation temporelle et spatiale.

L'algorithme proposé est basé sur l'estimation du vecteur de mouvement et de préservation maximale à priori des contours basée sur un modèle des champs de Markov. La figure ci-dessous montre le schéma de l'algorithme proposé.

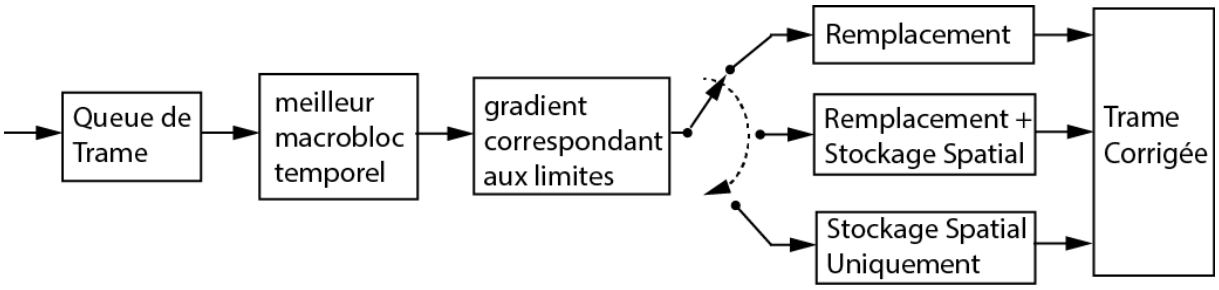


Figure 5.6 Schéma bloc de l'algorithme proposé dans [5].

Une mémoire tampon stocke les trames de référence pour la prédiction. Tout d'abord, la mise en correspondance de la limite classique est utilisée pour estimer le vecteur de mouvement; ensuite le mode de restauration est choisi (à moins de trois modes, à savoir uniquement temporel, spatio-temporelle ou uniquement spatial) sur la base de la valeur de la métrique GBM (Gradient-based Boundary Matching) d'appariement des limites à base du gradient.

La sélection du mode est basée sur une métrique GBM, destiné à rejeter la discontinuité de bord directionnelle entre le macrobloc de remplacement et la partie environnante de la trame. Dans le principe de la métrique GBM, seules les limites supérieures et inférieures du macrobloc manquant sont considérées. Pour chaque pixel de ces limites, une clique c et son complément c' sont pris en compte de même que huit directions entre le pixel et chaque composant de sa clique.

En comparaison avec la sélection du mode de sélection du macrobloc et la reconstruction uniquement spatiale, les résultats expérimentaux ont montré que l'algorithme GBM fournit une reconstruction plus agréable en évitant les discontinuités provoquées par le macrobloc manquant et son remplacement.

5.3.2 Un Algorithme avancé de dissimulation d'Artefacts pour des intra-trames dans le décodage H.264/AVC

Ce travail de recherche (voir [61]) fournit un algorithme de dissimulation d'erreur intra-trames dans le décodage H.264/AVC. L'algorithme proposé vise à effectuer la robustesse de la conception ayant le plus bas niveau de complexité possible. Cette technique de dissimulation n'augmente pas le débit, ne nécessite pas de modifications à l'encodeur, et n'introduit pas de retard.

La méthode propose un algorithme de dissimulation spatial basé sur l'étalement pondéré de pixel. Dans les cas de l'inter-trame, la méthode estime le mouvement dans le MB par les systèmes de prédiction à partir d'informations de mouvement spatiales et temporelles disponibles sur les sous-blocs voisins. La méthode élaborée consiste en trois principales approches décrites comme suit:

Tout d'abord, la corrélation temporelle entre trames (images) successives est explorée, ensuite des I-trames forcés sont introduits dans le flux binaire pour empêcher la propagation d'erreur dans des réseaux propices aux erreurs. Deuxièmement, un contrôle est effectué pour déterminer si un traitement temporel est possible avant d'essayer de corriger les blocs spatialement. La méthode développe une solution temporelle pour remplacer un MB manquant. Tout d'abord, pour chaque MB manquant, les sous-blocs 8x8 voisins sont utilisés pour déterminer un sous-bloc correspondant dans l'image précédente, comme indiqué ci-dessous.

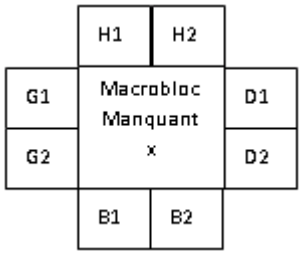


Figure 5.7 Les 8 sous-blocs adjacents au MB manquant

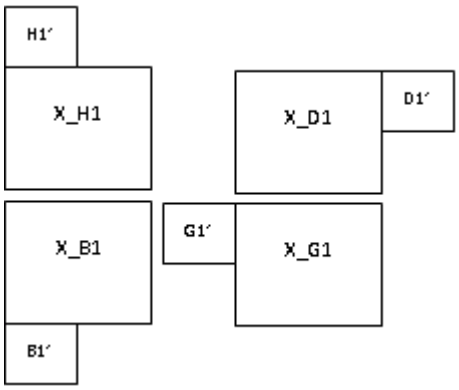


Figure 5.8 Sous-blocs correspondant dans la trame, et les MB candidats qui lui sont connectés

La somme des différences absolues est utilisée comme mesure de la similitude. Le MB manquant est remplacé par un MB estimé avec la somme des erreurs au carré des frontières la plus petite possible et son bloc le plus proche parmi les deux sous-blocs voisins de haut faisant partie des 8 sous-blocs voisins considérés. Enfin, un test est effectué pour vérifier la ressemblance avec le MB manquant.

Des expériences ont été effectuées qui montrent que l'algorithme proposé nécessite beaucoup moins de calculs et est plus robuste par rapport aux autres algorithmes de correction d'erreurs connus dans le décodage H.264/AVC.

5.3.3 Algorithme de dissimulation d'erreur basé sur le décodage H.264/AVC non-normatif

C'est une étude (voir [60]) qui propose une nouvelle méthode d'amélioration de la correction d'erreurs. L'idée principale est que les corrélations spatiales et temporelles sont utilisés conjointement pour la correction intra-trame et inter-trames. Le projet de correction d'erreurs comprend trois étapes principales: la détection de changement de scène, la détection d'activité de mouvement et la récupération du vecteur de mouvement.

Dans un premier temps, la méthode de correction d'erreurs utilise son algorithme de détection de changement de scène pour décider si le changement de scène se produit ou pas, puis il décide alors quel type d'information temporelle sera utilisé avec un algorithme de détection d'activité de mouvement approprié. Enfin, les vecteurs de mouvement sont dérivés de l'algorithme de récupération du vecteur de mouvement proposé.

Les résultats expérimentaux de la Figure 5.9 et la Figure 5.10 montrent de toute évidence que cette méthode de correction d'erreur est toujours meilleure que le décodage H.264/AVC non-normatif.

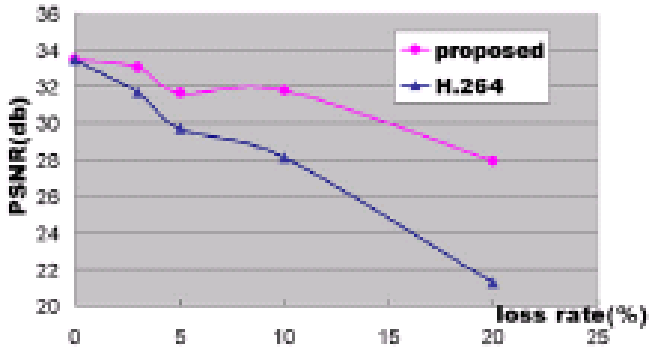


Figure 5.9 PSNR moyen vs. PLR (taux de perte de paquets) pour la séquence vidéo container avec une faible mobilité. Graphe tiré de [60].

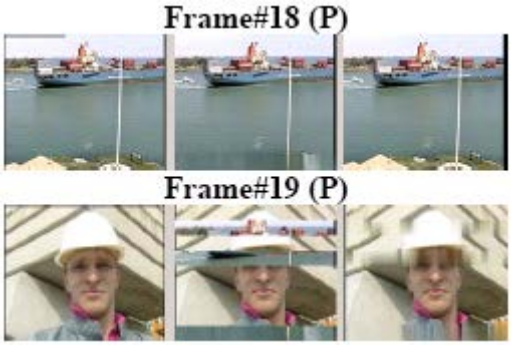


Figure 5.10 Trame décodée. Gauche: sans perte de paquets, Milieu :décodée avec 20% de PLR par H.264 non normatif ; Droite :décodée avec 20% par l'algorithme proposé en [60].

5.3.4 Elimination des effets de bloc dans les Systèmes de Décodage Vidéo basés sur les blocs

Cette technique propose une méthode pour surmonter le problème de perte de paquets et d'artefacts sur les réseaux sujets à des erreurs dans les systèmes de décodage basé sur des blocs. C'est une méthode spatio-temporelle 3D de dissimulation d'effets de bloc qui consiste premièrement à estimer un vecteur de mouvement pour le macrobloc perdu (ou défectueux) comme le montre la Figure 5.11 ci-dessous.

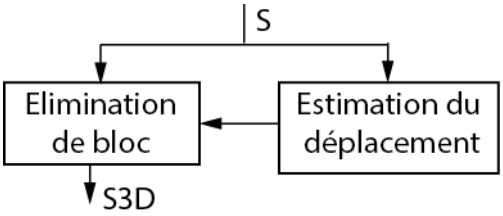


Figure 5.11 Diagramme bloc de la dissimulation 3D de l'effet de bloc.

Une unité de dissimulation des effets de bloc utilise le vecteur estimé pour la reconstitution temporelle et en outre pour la recherche spatiale d'échantillons de macroblocs voisins à la trame à corriger.

L'estimation de mouvement est faite grâce à une fonction de pondération w, qui est mise à 1 si l'échantillon d'image dans la trame courante est correctement reçue et à 0 sinon et une somme pondérée est associée à chacun des deux types d'images.

La méthode utilise en outre un filtre de dissimulation des effets de bloc pour réduire les artefacts de bloc lorsqu'ils apparaissent en raison d'occlusion et découverte d'objets. Ce filtre calcule deux gradients, représentant la somme des différences absolues le long d'une frontière de macroblocs et la valeur moyenne des deux sommes des différences absolues entre macroblocs voisins, puis les compare à deux seuils préalablement définis. Le système décide alors si la frontière de macroblocs contient des artefacts et si un filtre de dissimulation des effets de bloc est nécessaire.

5.3.5 Technique de convolution ou de déconvolution avec une gaussienne

Le filtrage est la méthode de correction d'erreurs la plus ancienne et la plus basique dans le traitement d'images. Il consiste à faire une convolution de l'image avec un filtre constituant le spectre de fréquence largement atteint par l'artefact à dissimuler.

Elimination du bruit par filtrage

Considérons un signal f en 2 dimensions qui a été dégradée par un bruit blanc gaussien n. Le signal dégradé g qui en résulte peut être exprimé comme suit:

$$g(x) = f(x) + n(x) \tag{5.1}$$

où x = (x,y). Le but de la dissimulation du bruit par filtrage est de supprimer n et d'extraire f de g. Dans les techniques de filtrage spatial, une estimation de f est obtenue en appliquant un filtre local h à g:

$$f(x) = h(x,\varsigma) \times g(x) \tag{5.2}$$

Dans une technique de filtrage spatial linéaire traditionnelle, le filtre local est défini sur la base de distances spatiales entre un point particulier (x,y) sur l'image et ses voisins. Dans le cas de la gaussienne le filtre local est défini comme suit:

$$h(x,\varsigma) = e^{-(1/2)(\|\underline{x}-\varsigma\|/\sigma)^2} \tag{5.2}$$

Dans l'équation (5.2), (x,ς) représente un point voisin à (x, y). De tels filtres fonctionnent sous l'hypothèse que la variation d'amplitude au sein d'un voisinage est assez petite et que le signal de bruit se compose de grandes variations d'amplitude. En lissant le signal sur un voisinage local, le signal de bruit devrait être supprimé dans cette hypothèse. Le problème avec cette hypothèse est que le détail du signal significatif est également caractérisé par de grandes variations d'amplitude. Par conséquent, ces filtres font apparaître des effets flous indésirables. Une solution simple et efficace pour ce problème est l'utilisation du filtrage bilatéral, à l'origine introduit par Tomasi et al. dans [62] et montré par Elad [21] comme émergeant de l'approche bayésienne.

Elimination du flou par déconvolution

L'algorithme de dissimulation de flou par déconvolution peut être utilisé efficacement en l'absence d'informations sur la distorsion (flou et bruit). Cet algorithme restaure la fonction d'étalement de points PSF (point-spread function) en même temps que l'image. L'algorithme de Richardson-Lucy amorti, accéléré est utilisé à chaque itération. Les caractéristiques d'un système optique supplémentaire (par exemple caméra) peuvent être utilisées comme paramètres d'entrée pouvant aider à améliorer la qualité de la restauration de l'image. Les contraintes PSF peuvent être transmises à travers une fonction définie par l'utilisateur.

Figure 5.12 Image CAMERAMAN: gauche (a) image bruitée et droite (b) restauré par filtrage bilateral. Obtenu de [44].

Voici la simulation d'une image de la vie réelle Im qui pourrait être floue (par exemple, en raison du mouvement de la caméra ou du manque de concentration): L'exemple simule le flou par convolution d'un filtre gaussien avec la vraie image. Le filtre gaussien représente alors une fonction PSF d'étalement de points. On aboutit à l'image floutée ImFlou (voir eq. (5.3)).

$$ImFlou = Im*PSF \tag{5.3}$$

Voici trois opérations de dissimulation de flou utilisant trois filtres PSF différents pour illustrer l'importance de connaître la taille de la vraie fonction PSF.

La première restauration, J1 et P1, utilise une matrice trop petite, SOUSPSF, pour une estimation initiale de la PSF. La taille de la matrice SOUSPSF est de 4 pixels plus courte dans chaque dimension que la vraie PSF.

La seconde restauration, J2 et P2, utilise une matrice de «1», SURPSF, pour une période initiale PSF qui est de 4 pixels plus grande dans chaque dimension que le vrai PSF.

La troisième restauration, J3 et P3, utilise un ensemble de «1», INITPSF, pour une première PSF qui est exactement de la même taille que la vraie PSF.

Figure 5.13 Image CAMERAMAN : gauche (a) : image originale et droite (b) : image floutée par convolution avec PSF. Obtenu de [62].

Toutes les trois restaurations produisent une PSF. La Figure 5.15 montre comment l'analyse de la PSF reconstruite pourrait aider à deviner la bonne taille pour la première PSF. Dans la vraie PSF qui estime un filtre gaussien, les valeurs maximales sont au centre (blanc) et diminuent aux frontières (noir).

La PSF P1 reconstruite dans la première restauration évidemment ne rentre pas dans la taille limitée. Elle a une variation de signal fort aux frontières. L'image correspondante, J1, ne montre aucune clarté améliorée par rapport à l'image floue, ImFlou.

Figure 5.14 Images restaurées J1, J2 et J3 respectivement avec les filtres suivants : de droite à gauche : SOUSPSF; SURPSF ; et INITPSF. Obtenu de [44].

La PSF P2 reconstruite dans la seconde restauration devient très lisse sur les bords. Cela implique que la restauration peut gérer un PSF d'une taille plus petite. L'image correspondante, J2, montre quelques déconvolutions mais elle est fortement endommagée par les suroscillations.

Enfin, la PSF P3 reconstruite dans la troisième restauration est un peu intermédiaire entre P1 et P2. La matrice, P3, ressemble d'ailleurs beaucoup à la vraie PSF. L'image J3 correspondante montre une amélioration significative, mais elle est encore corrompue par les suroscillations.

Les suroscillations de l'image restaurée, J3, se produisent le long des zones de contraste d'intensité forte et le long des bords de l'image. L'exemple suivant montre comment réduire ces effets en spécifiant une fonction de pondération. L'algorithme pondère chaque pixel selon la matrice de poids INITPSF tout en restaurant l'image et le PSF. Dans cet exemple, les pixels "pointus" sont d'abord évalués en utilisant la fonction de bord. Le seuil souhaitable est ensuite déterminé (ici égal à 0,3). Ainsi, toutes les valeurs de poids inférieur à 0.3 sont ignorées et celles supérieures ou égales à ce seuil sont assignées à 1.

L'image des bords est calculée par la fonction POIDS. Cette fonction prend en entrée une image Im en format de gris ou binaire en entrée, et renvoie une image binaire PB de la même taille que Im, avec des «1» là où la fonction trouve des bords sur Im et des «0» ailleurs. Pour élargir la zone, une fonction de dilatation des pixels à travers toute l'image peut être utilisée et une normalisation peut être effectuée. Les pixels proches des frontières sont également affectés à la valeur 0.

L'image est restaurée en faisant une déconvolution «aveugle» avec la matrice POIDS des poids et une augmentation du nombre d'itérations (30). La quasi-totalité des suroscillations est supprimée (voir Figure 5.17).

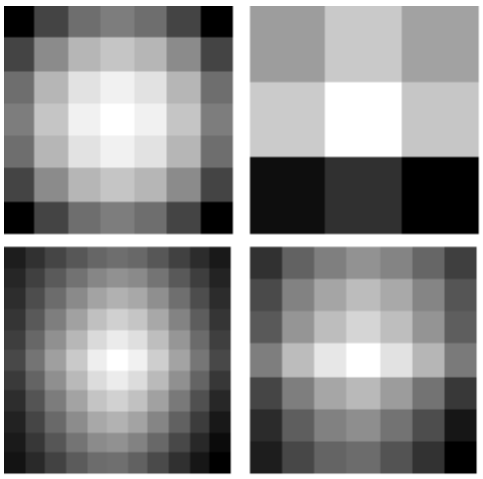


Figure 5.15 Aperçue image des filtres de répartition : de haut en bas et de gauche à droite PSF, SOUSPSF, SURPSF, INITPSF. Obtenu de [44].

Figure 5.16 Bords de l'image CAMERAMAN et image restaurée : à gauche bords de l'image; et à droite image restaurée avec réduction des suroscillations. Obtenu de [44].

5.4 Conclusion

La liste des types d'artefacts vidéo est non exhaustive. Dans ce chapitre les généralités sur les artefacts visuelles on été présentées, notamment les principales artéfacts visuelles présents dans le décodage de la famille des décodeurs vidéos MPEG. Ce sont le bruit, le flou, les effets de bloc, les suroscillations. Les sources de ces artefacts ont aussi été brièvement étudiées.

Il existe de nombreuses techniques de dissimulation d'artefacts dans l'état de l'art. Quelques techniques de dissimulation d'artefacts visuelles on été décrites. La dissimulation spatio-temporelle avec un mode de sélection perceptivement optimisé est d'une grande efficacité notamment sur la dissimulation des effets de bloc. L'algorithme élaboré en [61] pour la dissimulation d'artefacts sur des intra-trames dans le décodage H.264/AVC a montré des résultats satisfaisants. On a vu qu'une convolution ou une déconvolution peut permettre d'effectuer un filtrage du bruit ou de réduire le flou.

Une combinaison de certaines de ces techniques donnerait lieu à une approche générale de dissimulations d'artefacts visuels dans un décodeur. Ainsi, le filtrage peut être combiné à la méthode de dissimulation d'artefacts décrite en [60] pour une meilleure correction des erreurs dues à la famille des décodeurs MPEG.

6 Boucle de contrôle de la qualité du décodage

6.1 Introduction

Dans le chapitre précédent, des techniques avancées de corrections d'erreurs dans le décodage MPEG ont été étudiés. Dans les quatre premiers chapitres, des techniques pour l'évaluation de la qualité vidéo (ou d'une image au sens restreint aux aspects spatiaux) en utilisant la méthode avancée d'analyse statistique et l'intelligence artificielle ont été élaborées. Dans ce dernier chapitre, nous utiliserons les résultats de ces travaux pour aider à améliorer la qualité visuelle des images décodées par les décodeurs numériques de la famille MPEG. Pour ce faire, une boucle de contrôle intégrera une partie de correction des erreurs et une partie de mesure et de validation de la qualité de ces images. En effet, une fois la correction des erreurs effectuées sur une image donnée, il est nécessaire d'établir un seuil de validation du nouveau niveau de qualité de cette image corrigée afin de décider de valider sa qualité comme suffisante (selon l'œil humain) ou devant repasser par l'étape de correction d'erreurs. Ainsi on s'assure que les images décodées ont un niveau de qualité satisfaisant.

Afin de réduire la complexité de ce projet, nous allons considérer uniquement les aspects spatiaux d'une séquence vidéo, de sorte que le travail décrit soit applicable à une image (qui peut alors être vu comme une seule image dans la séquence d'image constituant une vidéo donnée). On désignera par trame une telle image fraisant partie d'une séquence d'images constituant une vidéo. Dans les projets futurs, nous prévoyons d'étendre ces expériences à une séquence vidéo tout en incluant des aspects de dissimulation d'artefacts vidéo liés au domaine temporel.

La validation est l'étape déterminante de ce chapitre consacrée à la boucle de contrôle de la qualité des images. Car un seuil de tolérance bien défini (par rapport au système visuel humain) assure un bon jugement à la sortie des images, quant à leur qualité visuelle. Il importe donc de définir au préalable la méthode de mesure de la qualité retenue, puis de fixer le seuil de qualité jugé adéquat. Toutefois, il n'existe en général pas de seuil standard pour la qualité des images. Ainsi, une étude avec des résultats statistiques expérimentaux permettra de montrer clairement les raisons du choix de notre seuil de qualité.

Le reste du chapitre se présente comme suit : Le paragraphe 6.2 présente l'analyse et le paramétrage des erreurs d'une image du décodage MPEG; La priorisation d'artefacts et la correction d'erreurs assistée par l'OMQV fera l'objet du paragraphe 6.3; Le paragraphe 6.4 présente l'étape de test de qualité suite à la mesure de la qualité d'une vidéo; le paragraphe 6.5 fait une comparaison avec un système commercialisé et présente des perspectives pour de futures projets et le paragraphe 6.6 conclut le chapitre.

6.2 Analyse et Paramétrage des erreurs du décodage MPEG

Lors du décodage d'une image donnée, le système de décodage doit contenir des algorithmes capables de détecter, de diagnostiquer et de corriger les principales erreurs faisant apparaître des artefacts visuelles sur les images en sortie du décodeur.

6.2.1 Analyse des Artefacts visuels dans le décodage MPEG

L'algorithme de correction d'erreurs proposé dans ce chapitre effectue d'abord une détection de la présence éventuelle des principaux artefacts liés aux décodeurs de la famille MPEG (le bruit, le flou, l'effet de bloc, les suroscillations) puis établi une priorisation de ces artefacts afin de les dissimuler par ordre croissante d'impact visuelle de dégradation de ces artefacts sur l'image. Ces artefacts ainsi que leurs origines sont explicités dans l'annexe (se référer aux paragraphes A.2 et A.3). En effet l'idée mise en œuvre ici se base sur la monotonie existant entre la dégradation de l'image par un de ces artefacts et le niveau de qualité de cette image. Une fois l'artefact prioritaire identifié, sa dissimulation sera alors faite dépendant de sa nature.

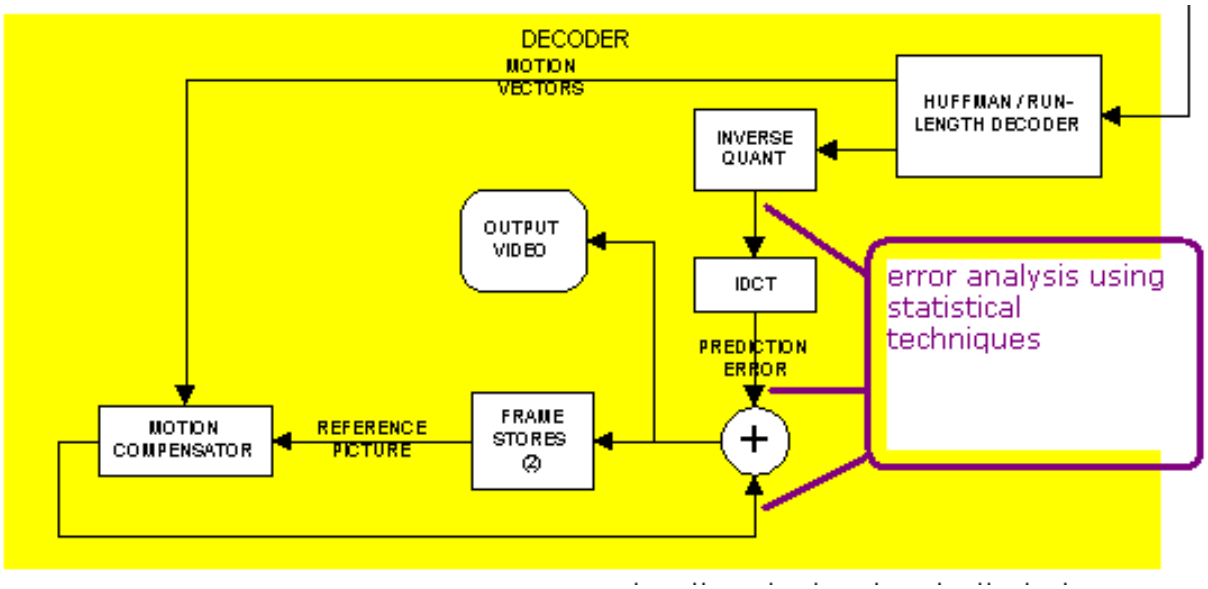


Figure 6.1 Schéma bloc illustrant l'analyse d'erreurs dans le décodeur MPEG.

L'analyse ici consiste à simuler puis reproduire les types d'erreurs (voir Figure 6.1) existant dans une image. Pour ce faire, une étude synthétique permet de reconnaitre les quatre principaux types d'artefacts (bruit, flou, effet de bloc et des suroscillations) en créant des paramètres de distorsion (liés aux artefacts) et des métriques de qualité étroitement liées à ces paramètres. La Figure 6.2 résume le processus de détection puis d'analyse des artefacts.

La création de métriques doit être normalisée. C'est-à-dire, les distorsions de synthèse produites sur les images doivent pouvoir être indiscernables aux artefacts originaux qu'on retrouve dans le décodage de la famille MPEG. Le choix des métriques et des paramètres a été fait de sorte que chacun des 4 artefacts de synthèse soit fortement corrélé à une métrique

donnée. Cette adéquation permettra alors d'avoir une sorte de discrimination entre les types d'artefacts lors de la priorisation par l'utilisation des coefficients de Spearman entre la métrique et la mesure de la qualité.

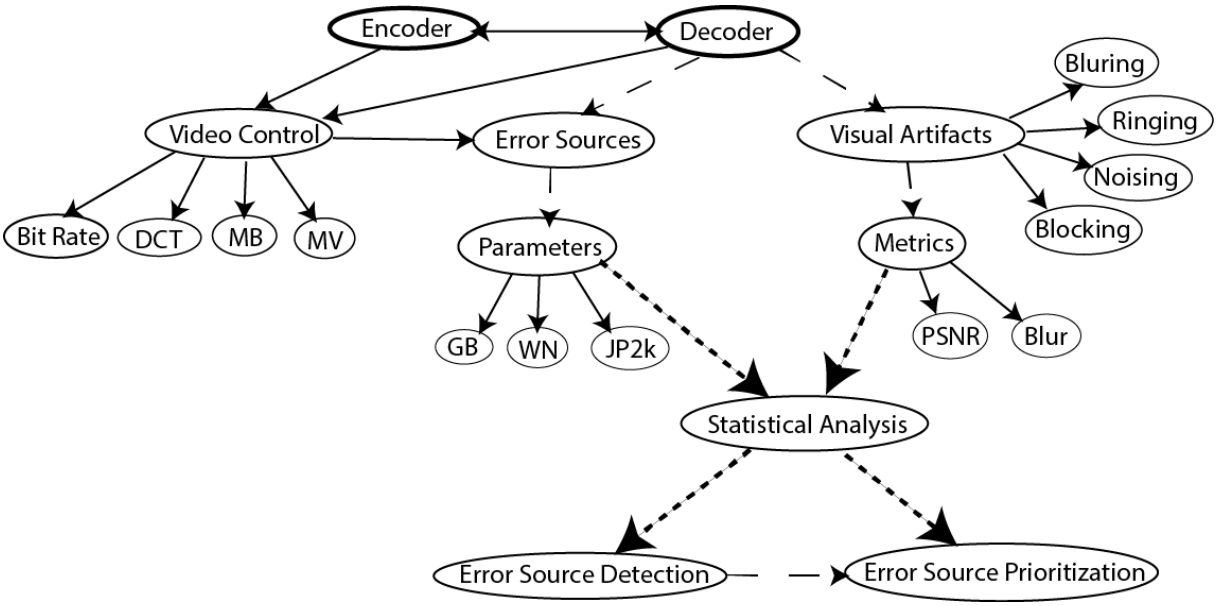


Figure 6.2 Détection et Analyse des erreurs dans le décodage MPEG.

6.2.2 Paramétrage de l'artefact "bruit"

Selon le type de bruit (bruit gaussien, bruit blanc, bruit additif, …) souhaitée il existe nombreux paramètres faisant apparaitre des bruit de synthèse dans une image.

Dans l'article [38], Li et Al crée synthétiquement les trois types de bruit cités en 5.2.1 (bruit structurel, bruit aléatoire, bruit additif). La forme la plus simple de créer un bruit facile à paramétrer est le bruit additif gaussien. Cette méthode consiste à additionner le signal d'entrée à un signal généré de façon aléatoire. Pour une répartition du bruit sur toute l'image, on choisit une distribution gaussienne du signal aléatoire. En effet le bruit gaussien est obtenu en ajoutant à chaque pixel d'une image une variable aléatoire suivant une loi de probabilité gaussienne.

Une loi de distribution gaussienne d'écart-type σ et de moyenne μ se présente sous la forme :

$$G_{\sigma,\mu}(s) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(s-\mu)^2}{2\sigma^2}} \tag{6.1}$$

En augmentant l'écart-type σ, on dégrade d'avantage l'image par ajout de bruit.

Soit la transformation de Box-Muller suivante :

$$\begin{cases} y_1 = \sqrt{-2\ln(x_1)}\cos(2\pi x_2) \\ y_2 = \sqrt{-2\ln(x_1)}\sin(2\pi x_2) \end{cases} \tag{6.2}$$

y1 et y2 suivent une loi normale centrée réduite si x1 et x2 suivent une loi uniforme. Comme le montre l'équation (6.2), le bruit additif Br généré sera produit à partir d'une

transformation de Box-Muller d'un signal aléatoire suivant une loi uniforme. L'équation de bruitage généralisée se traduit sous la forme de l'équation (6.3). Dans cette équation, la fonction y représente un bruit aléatoire (de la forme de y1 ou y2) suivant une loi gaussienne. Le bruit ajouté est donc créé à partir d'une multiplication de y avec le paramètre σ qui a été varié de manière à obtenir 3 différents niveaux de bruit (voir Figure 6.3). Ce bruit modélise bien un bruit additif réparti sur toute l'image.

$$ImBr(i,j) = Im(i,j) + Br(i,j) = Im(i,j) + \sigma.y(i,j) \quad (6.3)$$

Où Im est l'image originale sans artefacts;

ImBr est l'image bruitée obtenue après ajout du bruit blanc gaussien;

Br est le bruit blanc gaussien généré;

σ est l'écart-type du bruit blanc gaussien sur l'image ;

et y est une fonction aléatoire suivant une loi normale centrée et réduite

Figure 6.3 Une trame de la séquence mobile & calendar avec bruit blanc gaussien additif : de gauche à droite : originale tirée de [17]; bruité avec σ = 0,001 ; 0 ,01 et 0,1 respectivement.

Métrique liée au bruit

Bon nombre de métriques existent pour évaluer le niveau de bruit dans une image. Toutes sont basées sur le rapport signal sur bruit SNR (Signal to Noise Ratio). La plus simplifiée et la plus utilisée de ces métriques est le pic du rapport signal sur bruit PSNR (peak signal to noise ratio). Dans ce projet la métrique PSNR a été choisie pour la mesure du bruit. Se référer au chapitre 2 pour plus de détails sur le PSNR et les raisons de son choix.

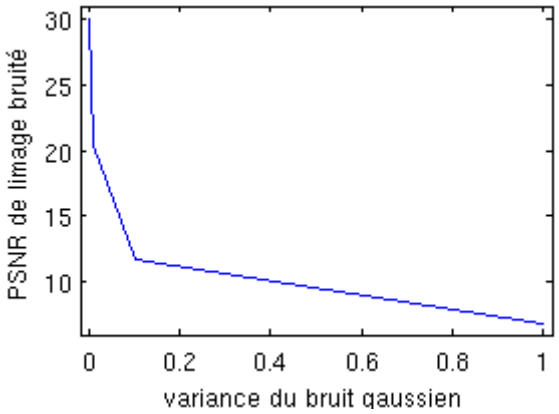


Figure 6.4 PSNR en fonction de la variance du bruit blanc gaussien.

6.2.3 Paramétrage de l'artefact "flou"

La convolution d'une image avec un filtre passe-bas – en fait un filtre gaussien à réponse impulsionnelle fini - est la méthode la plus utilisée en traitement d'image pour simuler le flou gaussien. Cette technique est celle choisi dans ce travail. Le paramètre faisant varier le flou est le facteur BT se définissant comme le produit entre la largeur B de la bande passante du signal fréquentiel à -3 dB et sa période T.

Dans la pratique, la génération de flou a été effectuée sur Matlab par les fonctions gaussfir et imfilter qui opèrent de la façon suivante : elle crée un filtre passe-bas gaussien à réponse impulsionnelle fini –ayant les caractéristiques :

BT : le produit de la période et de la largeur de bande à -3 dB. B est la bande passante unilatérale en Hertz ; T la période, exprimée en secondes.

NT : le nombre de périodes de symbole entre le début de la réponse impulsionnelle du filtre et le point maximum du signal.

OF : le facteur de sur-échantillonnage, c'est le nombre d'échantillons par symbole. Si il n'est pas spécifié, DE = 2 est utilisé.

Cette fonction effectue une convolution entre le filtre PSF (cf. section 5.3.5 p. 99) et l'image d'entrée Im pour obtenir l'image floutée ImFlou. En augmentant BT, on diminue le niveau de flou et inversement comme le montre la Figure 6.5.

Figure 6.5 Une trame de la séquence foreman avec flou gaussien: de gauche à droite : originale tirée de [17]; floutée avec BT = 0,1 ; 0 ,06 et 0,08 respectivement.

Métrique liée au flou

La métrique utilisée dans ce projet pour la mesure du flou a été établie par F. Crete et Al. Cette métrique est définie au paragraphe 2.3.2 parmi les métriques utilisées dans le projet.

Pour une simplification de la réalisation de nos travaux, les seuls artefacts pris en compte étaient les artefacts "bruit" et "flou" paramétrés ci-dessous. Toutefois dans le reste de ce paragraphe, nous présentons une étude du paramétrage des deux autres artefacts, à savoir les effets de bloc et les suroscillations.

6.2.4 Paramétrage de l'artefact "effets de bloc"

Très peu de métriques connues dans l'état de l'art illustrent convenablement des artefacts d'effets de bloc dans le domaine du décodage MPEG. Une métrique sans référence élaborée par Wang en [66] a retenue notre attention par son indépendance totale vis-à-vis de la référence. Mais elle ne représente pas au mieux les effets de bloc rencontrés dans le décodage MPEG. L'étude expérimentale [23] propose une façon synthétique de créer et de paramétrer les effets de bloc. L'équation (6.4) illustre le résultat obtenu par l'algorithme proposé. Cette méthode de synthèse de paramétrage des effets de bloc nous a particulièrement intéressée parce qu'elle représente très exactement le comportement réel des effets de bloc créés dans la famille

des décodeurs MPEG. La Figure 6.7 présente le résultat de génération par cette méthode des effets de bloc sur une image.

$$ImBloc\ (i,j) = Im\ (i,j) + n.D\ (i,j) \quad (6.4)$$

Où Im est la matrice représentant l'image d'origine;

i et j les positions spatiales du pixel sur l'image ;

n un nombre constant faisant varier le niveau d'importance des effets de bloc;

D est la différence entre la moyenne d'un bloc B1 de 8x8 pixels et la moyenne du bloc B2 de 24x24 pixels dont B1 est le centre;

et ImBloc est la matrice de l'image obtenu contenant les artefacts de bloc ainsi générés.

6.2.5 Paramétrage de l'artefact "suroscillations"

La méthode de synthèse des suroscillations décrite en [23] est particulièrement intéressante pour les mêmes raisons que précédemment, car elle reproduit les suroscillations conformément aux suroscillations réelles générées par le décodage MPEG. Cet algorithme de génération des suroscillations se compose d'une paire de filtres complémentaires liés par l'équation (6.5).

$$H(z) + G(z) = \rho.z^{-n_0} \quad (6.5)$$

Où H et G sont des fonctions de transfert de filtres passe-haut et passe-bas à réponse impulsionnelle finie respectivement.

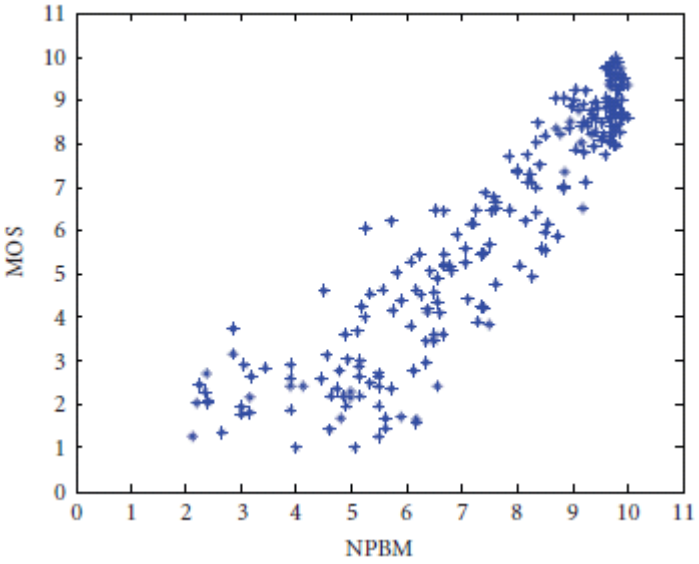


Figure 6.6 Correlation entre MOS et la métrique NPBM. Obtenu de [39].

Les paramètres ρ et n_0 permettent de varier l'intensité des artefacts. Pour ρ = 1 et n0 = 0, la transformée en z de la sortie de notre système est donnée par l'équation suivante (6.6). Cette équation illustre le résultat obtenu par l'algorithme proposé.

$$ImSon\ (z) = [H(z) + G(z)]\ .\ Im\ (z) \quad (6.6)$$

Figure 6.7 Artefacts générés dans une zone d'une trame de la frame N°320 de la vidéo «calendar » : à gauche l'image originale ; au centre l'image avec effets de bloc ; à droite l'image avec des suroscillations. Tiré de [23].

Le mode d'utilisation de l'OMQV conçu ici est le mode B illustré dans l'introduction du manuscrit (voir Figure 0.1). Dans le reste du manuscrit, nous ne considérerons que les artefacts relatifs au décodage de la famille MPEG [23-24]. Pour simplifier l'algorithme de correction et les démarches, dans la suite, seuls les artefacts visuels "flou" et "bruit" sont considérés, car ils sont les plus influents sur l'image. Les variations de score de qualité visuelle d'une image en fonction des mesures de niveau du flou et de bruit sont présentées dans les Figures 6.9 et 6.10).

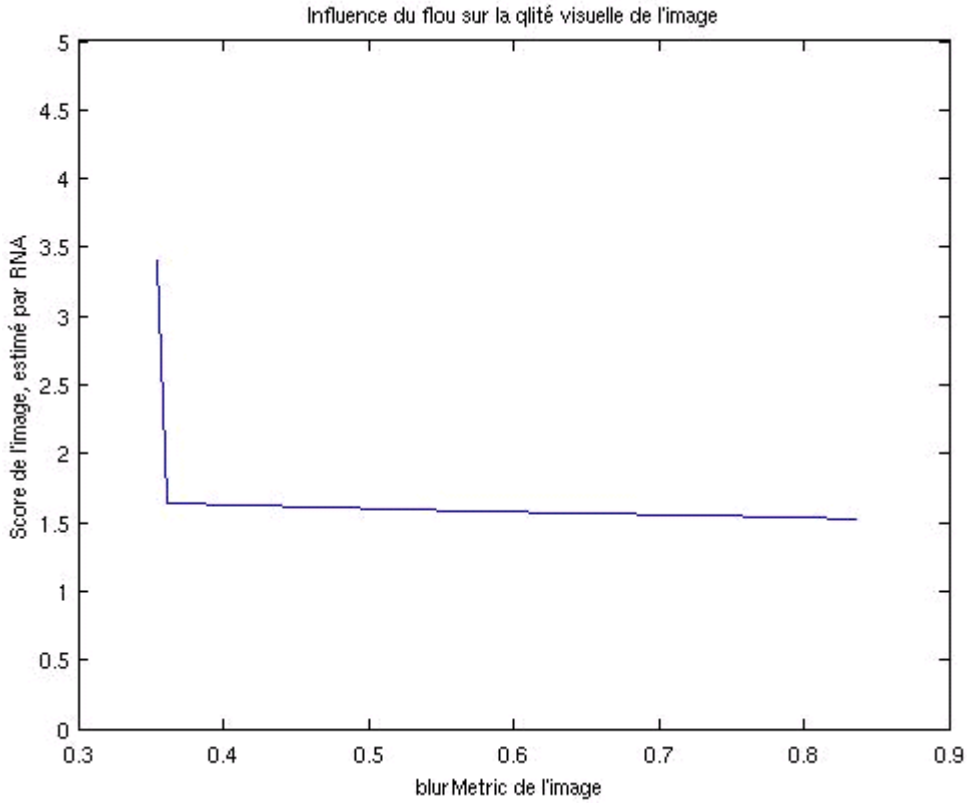


Figure 6.8 Variation du score de qualité visuelle en fonction du flou de l'image.

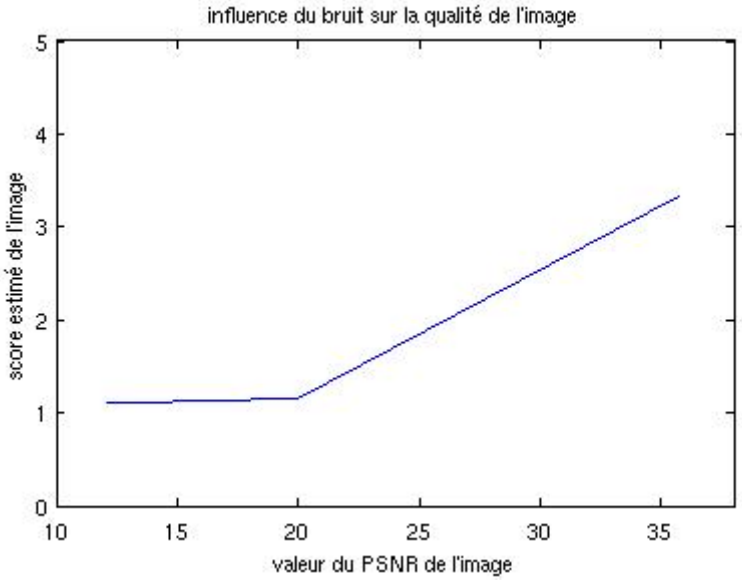


Figure 6.9 Variation du score de qualité visuelle en fonction du bruit de l'image.

6.3 Priorisation et Elimination d'artefacts visuels

Afin d'améliorer la qualité d'une image, l'algorithme de correction proposé dans ce chapitre détermine l'artefact qui l'influence le plus. L'idée de cette priorisation consiste dans un premier temps à calculer les coefficients de Spearman entre chacune des 2 métriques énoncées dans la section précédentes - mesurant les variations de chacun des 2 principaux artefacts – et la moyenne MOS des scores donnés par les observateurs. Par la suite l'algorithme considérera comme «prioritaire» l'artefact ayant le plus haut coefficient de Spearman avec le MOS. Cette idée suppose deux hypothèses : L'une, une monotonie entre chacune des 2 métriques représentant les niveaux de distorsions suivant les 2 artefacts et la moyenne MOS des scores de qualité donnés par les observateurs ; et l'autre suppose de très faibles corrélations croisées entre les métriques des artefacts, compte tenu des influences croisées pouvant exister entre les sources de ces artefacts.

6.3.1 Priorisation des artefacts suivant leur degré d'impact sur la qualité des images

La Figure 6.10 présente l'identification de la distorsion MAD (*most annoying distorsion*) ayant le plus grand impact sur la détérioration de la qualité de l'image. Le coefficient de Spearman examine s'il existe une relation entre chaque métrique mi($1 \leq i \leq 4$) et la moyenne des scores observés MOS, sous l'hypothèse de l'existence de relations monotones. Ce coefficient est très utile lorsque l'analyse du nuage de point révèle une forme curviligne dans une relation qui semble mal s'ajuster à une droite.

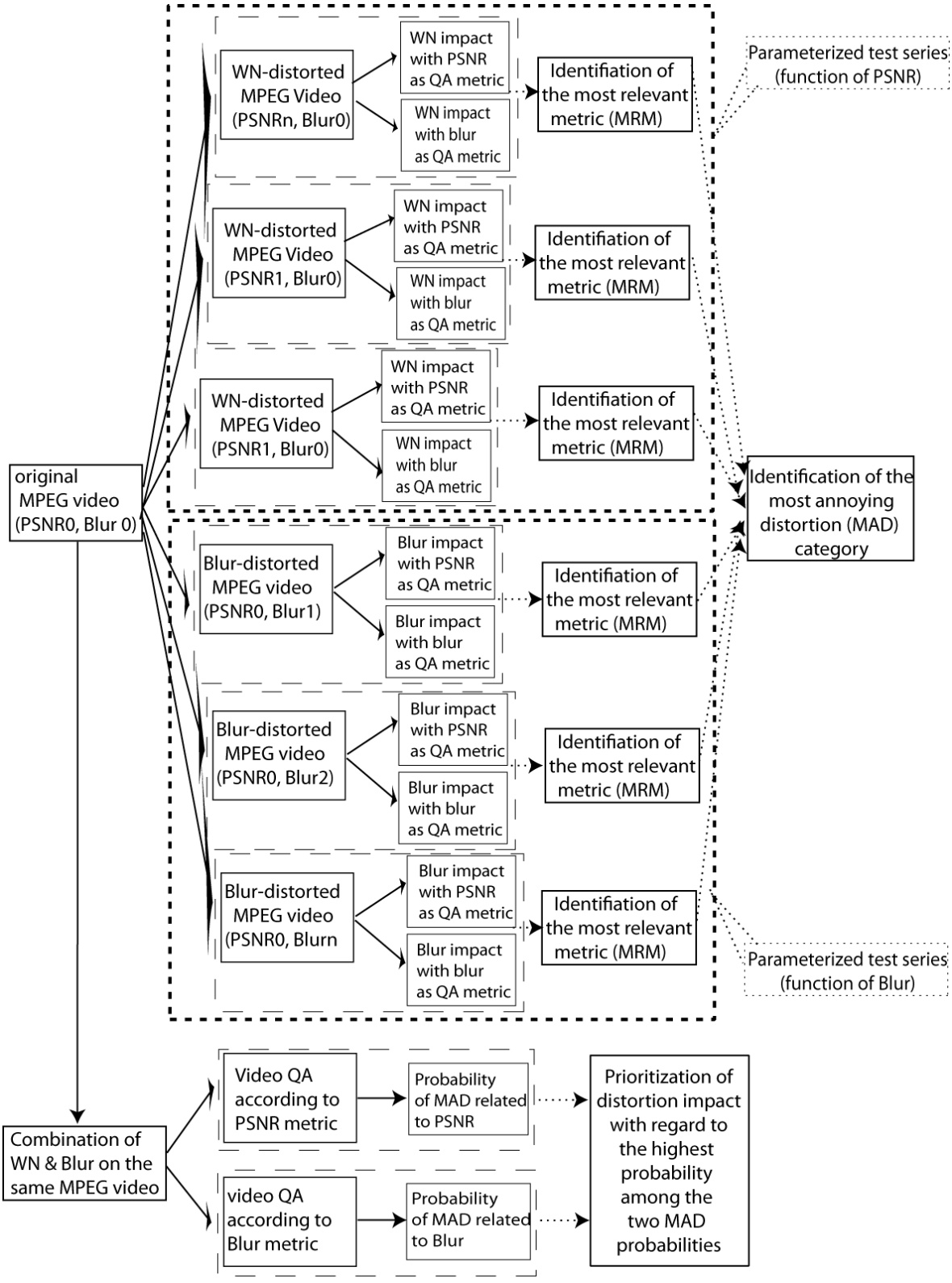


Figure 6.10 Graphe de priorisation des artefacts. Les 2 métriques *PSNR* et *blur* mesurent resp. le bruit et le flou, ne sont pas totalement indépendantes et sont monotones avec le MOS (voir Figure 6.9 et Figure 6.10). L'artefact (WN ou Blur) susceptible d'être le plus perceptuellement nuisible se mesure à la probabilité qu'une détérioration visuelle de la qualité (mesurée via le MOS) provienne d'une diminution de *PSNR* ou d'une augmentation de *blur*.

6.3.2 Algorithme de correction d'erreurs assisté par l'OMQV

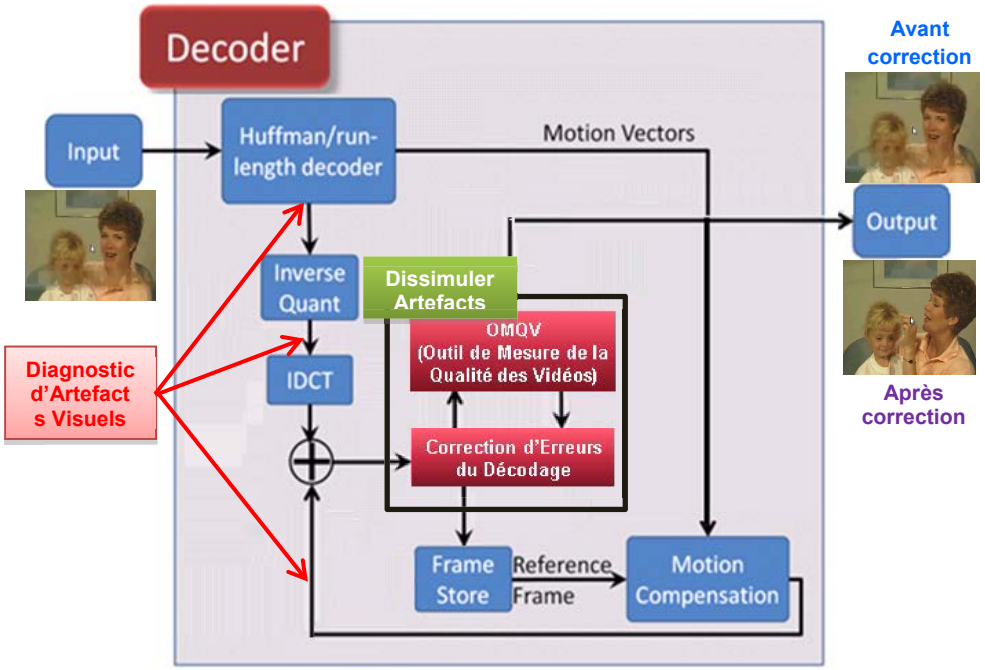


Figure 6.11 Correction d'erreurs assisté par l'OMQV dans un décodeur MPEG.

La Figure 6.11 représente le schéma de conception de l'outil de correction des artefacts dans un décodeur numérique. La méthode utilisée dans cette sous-section pour réaliser la correction d'artefacts assistée par l'OMQV est basée sur les réseaux de neurones artificiels (chapitre 3). Pour la correction des erreurs, la méthode utilisée sera l'algorithme du filtre adaptatif de Wiener car elle s'adapte mieux à la correction simultanée de flou et de bruit..

On considère donc une image dégradée $g = h * u + b$, où u est l'image originale à restaurer, h un noyau de convolution positif (réponde impulsionnelle du filtre « flou ») et b est un bruit de loi de probabilité (identique pour chaque pixel) μ. On prendra ici un bruit blanc gaussien d'écart-type σ_b.

Le flou, considéré gaussien, sera modélisé par une gaussienne de variance σ^2:

$$h(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (6.1)$$

Où x et y désignent les coordonnées en pixels d'un point de l'image ;

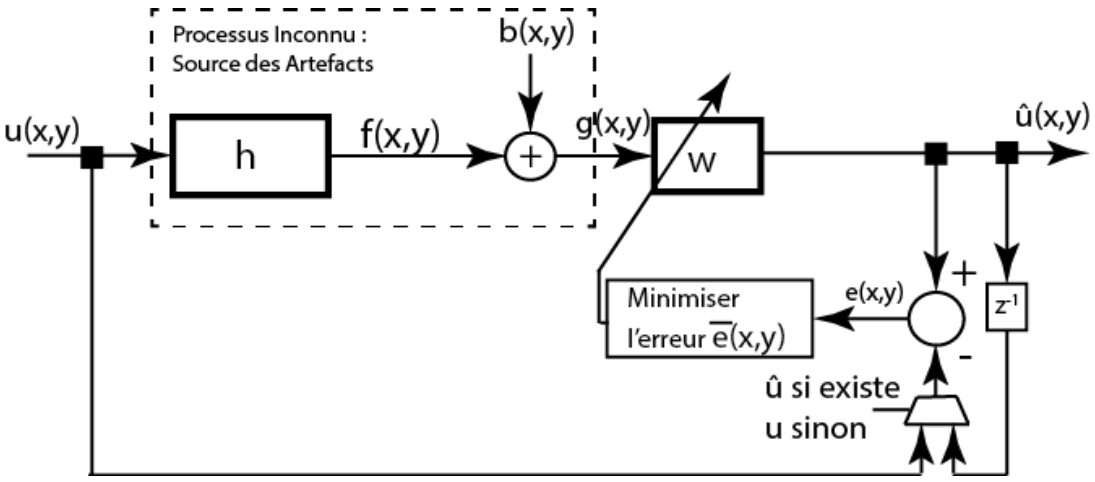


Figure 6.12 Schéma de la correction d'erreurs assistée par l'OMQV.

On a les équations suivantes pour décrire ce système:

$$\|e\|^2 = \|g - h * u + b\|_{L^2}^2 \qquad (6.2)$$
$$= \hat{g}^2 + (\hat{h}\hat{u} + \hat{b})^2 - 2\hat{g}(\hat{h}\hat{u} + \hat{b})$$

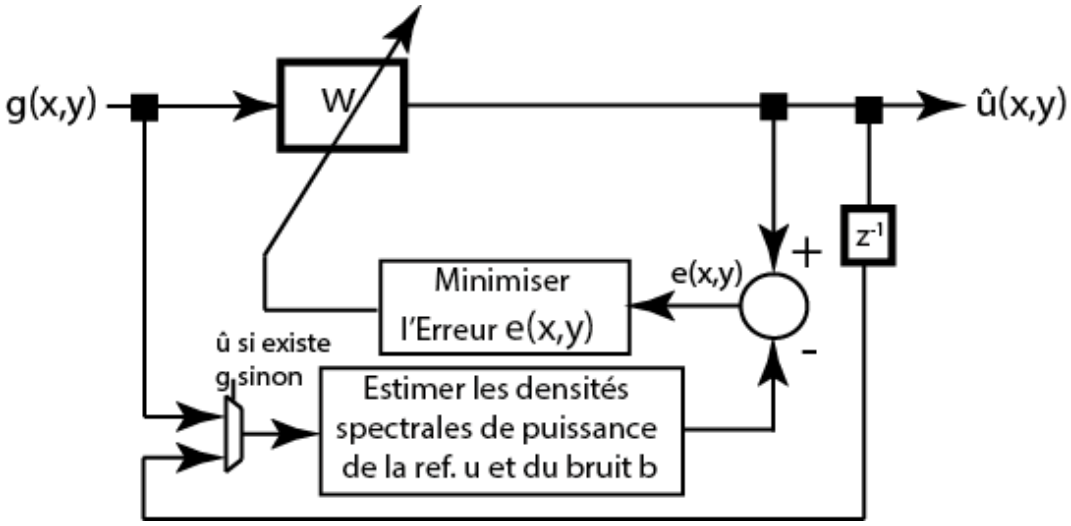


Figure 6.13 Schéma réel de la boucle de correction des images.

Recherche de la solution optimale $\hat{u}$:

Dans la pratique, l'image de référence u est inconnue, mais son spectre de puissance Pu est nécessaire dans l'algorithme de correction d'erreur et peut être estimée. Une estimation du spectre de puissance Pu de u est obtenue à partir de l'image d'entrée g, en supposant que les deux images ont des spectres de la même forme. Donc le problème en temps réel peut être illustré comme représenté sur la Figure 6.14

Imposer des contraintes d'optimisation dans la recherche de $\hat{u}$ permet de contrôler les paramètres du filtre de Wiener, σ_b(l'écart type de bruit gaussien) et N (la dimension de h), et donc permettre d'accroître la qualité de l'image optimale

$$g = h * u + b \tag{6.3}$$

Solution en l'absence de bruit:

Voyons la fonction de transfert de PSF du flou h.

Supposons que nous appliquons le filtre h à l'image d'entrée g pour obtenir une estimation $\hat{u}$ de u:

$$\hat{u} = h * g \tag{6.4}$$

L'énergie du bruit peut s'écrire:

$$|h * \hat{u} - g|^2 = |b|^2 \tag{6.5}$$

Une bonne solution doit satisfaire à l'équation (6.5). Une solution simple serait de supposer que l'énergie du bruit est trop faible et il suffira alors de choisir l'estimation $\hat{u}$ qui minimise le réel $|h * \hat{u} - g|^2$.

Après dérivation, il vient:

$$\frac{\partial (h * \hat{u} - g)^T (h * \hat{u} - g)}{\partial g} = 2h^T (h * \hat{u} - g) \tag{6.6}$$

En recherchant les racines de cette dérivée, nous obtenons l'extremum:

$$\hat{u} = h^{-1} * g \tag{6.7}$$

C'est le filtre inverse de h, la fonction de transfert de flou. Il admet une solution déterministe que si $|b|^2 = 0$, Mais cette solution est irréalisable pour le bruit dans les hautes fréquences. En d'autres termes, c'est le cas lorsque le seul artefact visuel est le flou. Ensuite, la Figure 6.13 est modifiée et présente désormais une boucle ouverte comme le montre la Figure 6.15 ci-dessous:

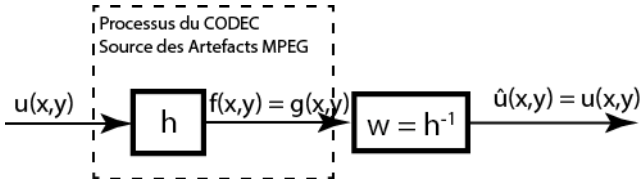


Figure 6.14 Correction d'erreurs dans le cas d'un bruit nul.

Dans ce cas, la taille *d* de la matrice PSF devient le seul paramètre à maîtriser dans le filtre de Wiener. Cela montre que le flou de l'image à corriger est plus influencé par *d* que parσ_b. La Figure 6.16 ne montre pas très clairement ces résultats parce que les variations sont très faibles. Mais une conjecture à un ordre plus élevé de *d* peut permettre de voir cette légère monotonie entre *d* et la métrique de flou blurMetric. La Figure 6.16 montre que le bruit reste quasiment identique sur l'image traitée, lorsqu'on fait varier la taille *d* de la matrice du masque de flou. D'autre part, ces deux Figures montrent également que la valeur du PSNR et blurMetric dépend de la parité de *d* et que ces valeurs sont optimales (dans le sens de la meilleure qualité de l'image traitée) lorsque *d* est impair. Cette dernière assertion peut être démontrée à partir de la base des équations (6.8) et (6.13). En fait, ces équations montrent que la valeur des pixels de

l'image traitée au niveau du bord de l'image dépend de la parité de *d*, et que la valeur optimale de 'h' est obtenue quand *d* est dans la forme *d* = 2p +1, avec p, un nombre entier de positif.

Cas général: la minimisation de MSE et l'optimisation en utilisant la fonction gradient:

Résolvons l'équation (10). L'objectif est d'estimer image de référence u par son estimation $\hat{u}$ en fonction de l'image d'entrée g. Nous nous proposons de trouver cette estimation par la méthode d'optimisation du minimum local de l'erreur quadratique moyenne LLMMSE (local linear minimum of MSE) sur la base (de tout ou une partie) de g. La version de filtre FIR de ce problème peut être explicite comme suit:

$$\hat{u}(x, y)= \sum_{i=n-d}^{n} \sum_{j=m-d}^{m} h_{x+i,y+j} * g(x,y) \qquad (6.8)$$

Où x et y sont les coordonnées du pixel dans l'image.

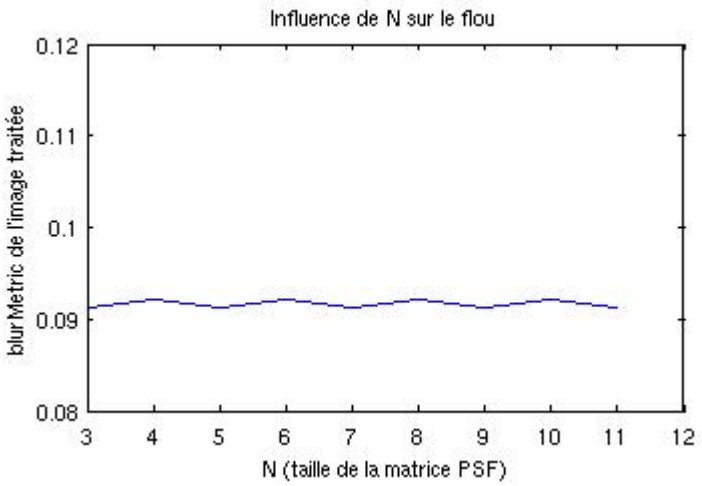


Figure 6.15 Variations du niveau de flou de l'image en fonction de la taille *d* du PSF en cas d'absence de bruit additif sur l'image : les minimas de niveau de flou sont atteints pour les valeurs de *d* impairs. Ils correspondent aux images de qualité visuelle optimale.

On pose $\hat{u}(x, y) = \alpha g(x, y) + \beta$

L'image et le bruit sont supposés indépendants. Nous considérons un bruit blanc gaussien additif de varianceσ_b

α et βsont choisis et de façon à minimiser l'erreur quadratique moyenne de e(x, y). Le gradient de ce nouveau problème peut alors être calculé comme suit:

$$J(\alpha, \beta) = E\left[(\widehat{u(x, y)} - u(x, y))^2\right] \qquad (6.9)$$
$$= E[(\alpha\, g(x, y) + \beta - u(x, y))^2]$$

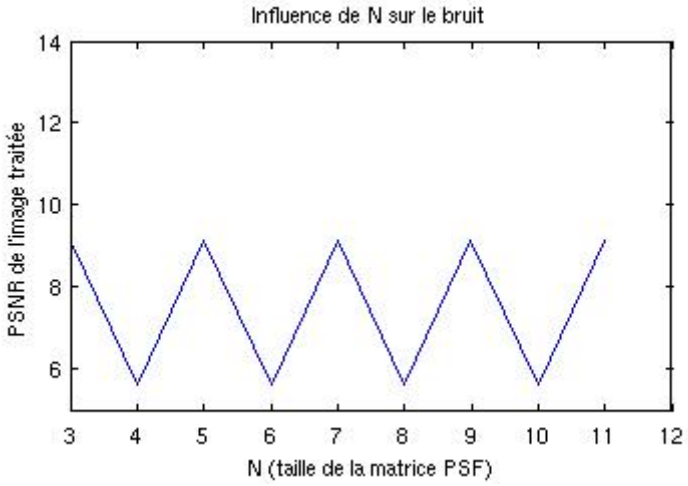


Figure 6.16 Variation du bruit de l'image en fonction de la taille *d* du masque PSF : les maximas de niveau de bruit sont atteints pour les valeurs de *d* impairs. Ils correspondent aux images de qualité visuelle optimale.

Après dérivation, on obtient:

$$\frac{\partial J(\alpha,\beta)}{\partial \alpha} = 2E[(\alpha g(x,y) + \beta - u(x,y))g(x,y)]$$
$$\frac{\partial J(\alpha,\beta)}{\partial \beta} = 2E[\alpha g(x,y) + \beta - u(x,y)] \quad (6.10)$$

On annule alors la dérivée pour trouver les extremums en utilisant le fait que $\overline{g(x,y)} = \overline{u(x,y)}$ et que $\sigma_g^2 = \sigma_u^2 + \sigma_b^2$, on obtient:

$$\alpha = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_b^2} \quad (6.11)$$

$$\beta = (1 - \alpha)\overline{g(x,y)}$$

L'estimation de l'image optimale est alors:

$$\widehat{u(x,y)} = \frac{\sigma_u^2}{\sigma_g^2} g(x,y) + \frac{\sigma_b^2}{\sigma_g^2}\overline{g(x,y)} \quad (6.12)$$

Le choix de *d* et σ_b^2 est très important. *d* doit être au moins égal à 3. Ces deux valeurs sont dépendantes de l'expression du filtre au sens LLMMSE, qui peut être interprétée comme une réponse locale ou un masque dont les poids $w_n (0 \le n < d)$ sont des fonctions de statistiques locales dans l'entrée.

$$"h_n" = \begin{cases} \alpha + \dfrac{1-\alpha}{d} \; if \; n = 0 \\ \dfrac{1-\alpha}{d} \; if \; |n| \le (d-1)/2 \\ 0 \; else \end{cases} \quad (6.13)$$

On peut aussi déterminer une réponse en fréquence locale dont la bande passante est une fonction du rapport signal sur bruit local.

Passons maintenant au domaine de fréquence: (x, y) sont remplacés par (u, v) et une lettre minuscule se référant à un signal ou une fonction dans le domaine spatial est maintenant écrit en majuscules l'image restituée est donnée par:

$$\widehat{U}(u,v) = W(u,v).G(u,v) \tag{6.14}$$

Le filtre de Wiener est choisi de façon à minimiser l'erreur quadratique moyenne$\mathrm{E}\left[\left|\widehat{\mathrm{U(u,v)}} - \mathrm{W(u,v).G(u,v)}\right|^2\right]$. Supposons que l'image est indépendante du bruit ; le filtre de Wiener est alors donné par:

$$W(u,v) = \frac{H(u,v)}{|H(u,v)|^2 + \frac{S_b(u,v)}{S_u(u,v)}} \tag{6.15}$$

Où $S_u(u,v)$ est la densité spectrale de puissance de l'image de référence U, et $S_b(u,v)$ est la densité spectrale de puissance du bruit B. Dans notre cas, pour un bruit additif blanc gaussien de variance σ_b^2 pour un pixel, et une image numérique de résolution mxn, nous avons:

$$S_b(u,v) = M \times N.\sigma_b^2$$

σ_b^2 est estimé par les écarts locaux de régions lisses de basse fréquence dans l'image. $S_u(u,v)$ peut être estimée en utilisant l'image de référence G (u,v), car le filtre de Wiener n'est pas sensibles à de faibles variations de spectre de puissance et les transformations sur une image ne font pas varier (ou font très peu varier) son spectre de puissance.

Correction d'erreurs assistée par l'OMQV: simulation et résultats

Première méthode: Les fonctions de coût pour mesurer l'optimisation de l'image sont les deux métriques d'artefacts blurMetric et PSNR

Pour une valeur fixe de la taille N de la matrice du masque h, (valeur initiale prise d = 3 et maximum d = 11), on a commencé en corrigeant le flou dans l'image lors de la première itération étant donné le fait que son influence sur la qualité visuelle (mesurée par les scores estimés par l'OMQV) comme on le voit dans le début du présent article, est supérieure à l'influence du bruit sur la qualité de l'image.

Nous avons prouvé dans le premier paragraphe que la mesure de flou blurMetric et le PSNR ont des dépendances monotones avec la qualité de l'image mesurée au travers des scores estimés par l'OMQV. En outre, ces deux mesures ont des dépendances sur le paramètre d (taille de la matrice du masque gaussien h). Mais ils ont aussi des dépendances monotones sur sigma, cela est prouvé sur les équations (6.11), (6.12) et (6.13) et ces dépendances sont représentées sur les Figures 6.18 et 6.19.

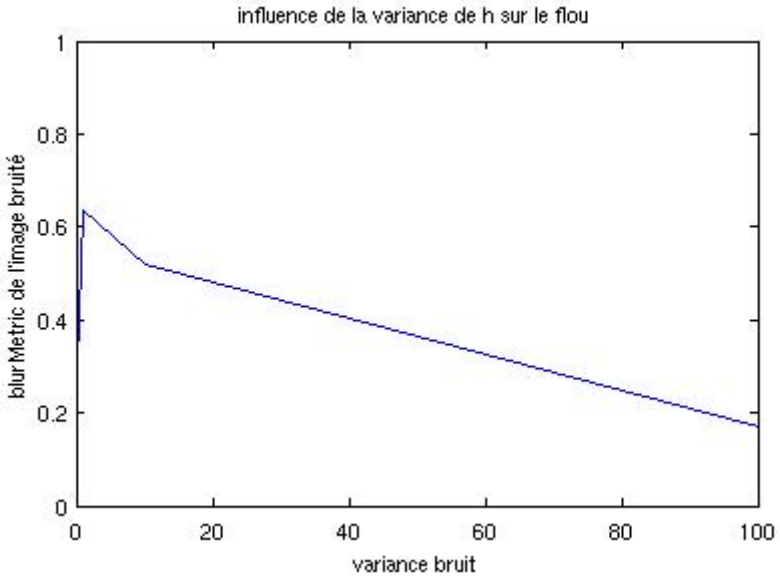


Figure 6.17 Variation du flou de l'image traitée avec l'écart-type σ_b du bruit gaussien.

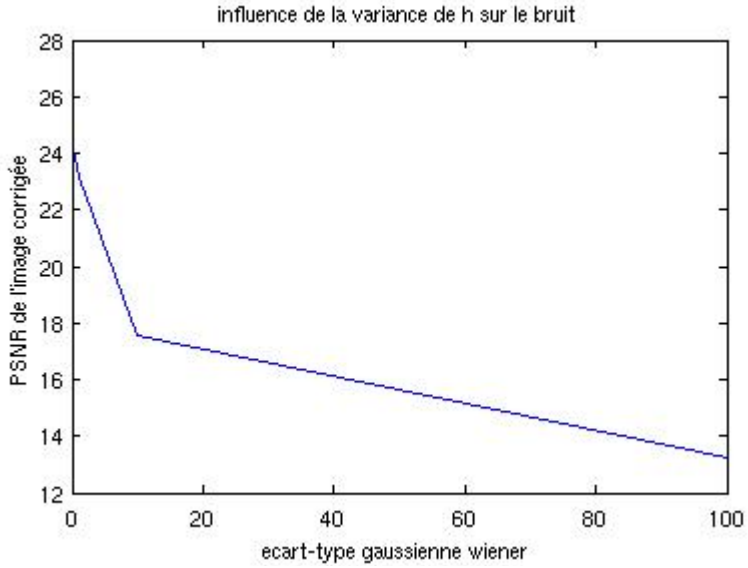


Figure 6.18 Variation de PSNR de l'image traitée avec l'écart-type σ_b du bruit gaussien.

Nous affirmons donc sur la base de ces hypothèses que nous pouvons utiliser les métriques mesurant le flou et le bruit comme fonctions de coût pour l'algorithme d'optimisation de la qualité visuelle des images. Les résultats de simulation de cette méthode sont représentés sur la Figure 6.20.

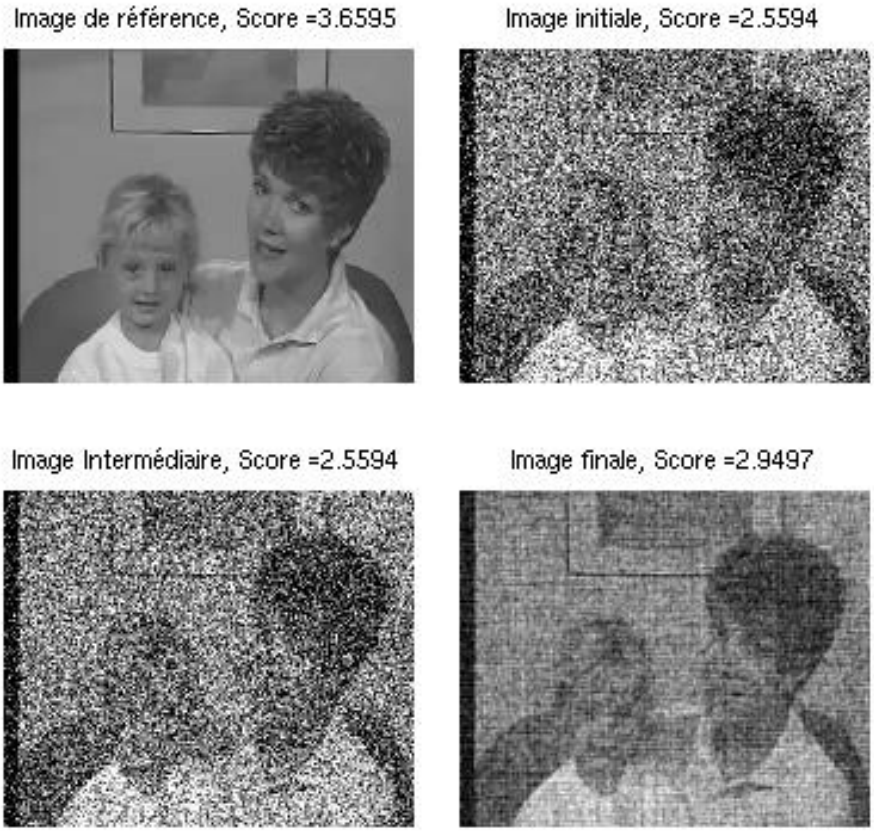


Figure 6.19 Résultats de simulation : mesures des artefacts prises comme fonctions coûts.

Le score n'a pas évolué entre l'image initiale et l'image intermédiaire. Ceci s'explique par le fait que la fonction coût considérée dans l'optimisation n'est pas le score mais les métriques. En revanche on voit une amélioration visuelle sur l'image finale, tout comme une amélioration du score de qualité. Ceci explique clairement que la monotonie entre le score de qualité et les métriques est globalement observée d'une part. Et d'autres parts, ce score de qualité estimé est bien une indication objective sur la qualité visuelle de l'image.

Deuxième méthode: La fonction de coût pour mesurer l'évolution de l'optimisation d'image est le score de qualité visuelle estimé par l'OMQV

Une estimation $\hat{u}$ de l'image de référence de u est calculé à chaque itération i et le score correspondant $score_i$est évalué par l'OMQV afin de mesurer le sens de l'évolution du processus de correction d'erreurs. Puis la monotonie vu entre la taille N du masque h et la métrique blurMetric mesurant le flou de l'image d'une part et entre cette même métrique de mesure du flou et le score de qualité visuelle de l'image estimé par l'OMQV d'autre part nous a permis de définir une valeur optimale Nopt de N, correspondant à l'itération i_{opt} correspondant à l'image traitée ayant le score de qualité le plus élevé. A partir de cette valeur Nopt, l'autre paramètre de détérioration de la qualité σ_b a été varié à partir d'une valeur initiale 0,01 à jusqu'à la valeur finale 100.

De la même manière, la monotonie observée entre l'écart-type σ_b et le PSNR de l'image d'une part et entre le PSNR et le score de qualité visuelle de l'image estimé par l'OMQV

d'autre part nous a permis d'obtenir le meilleur score $score_{opt}$. Ce score est le plus haut score possible de l'image traité, et l'image correspondante a été prise comme l'estimation optimale de l'image de référence u. Les résultats de simulation de cette méthode sont représentés sur la Figure 6.21.

Le score a évolué entre l'image initiale et l'image intermédiaire. Ceci s'explique par le fait que la fonction coût considérée dans l'optimisation est bien le score de qualité visuelle. On peut très bien remarquer une amélioration visuelle sur l'image finale et une amélioration du score de qualité.

La qualité visuelle de l'image finale n'est pas meilleure que celle de l'image de référence bien que le score de qualité obtenu est plus élevé. C'est une incohérence qui est dû à l'insuffisance de la représentativité da la variation de l'artefact flou dans la base de données que nous avons considérée. En effet, cette base de données n'avait pas du tout de représentativité de l'artefact flou. N'ayant pas eu assez de temps pour établir une base de données adéquate, nous avons essayé de résoudre ce problème en ajoutant aux images extraites des vidéos de la base de données d'autres images créées à partir des images d'origine par ajout de flou gaussien. Une amélioration des résultats pourrait être obtenue si la représentation du flou était égale à celle des autres paramètres (PLR, PSNR, SI), uniformément représentées sur toute la base de données.

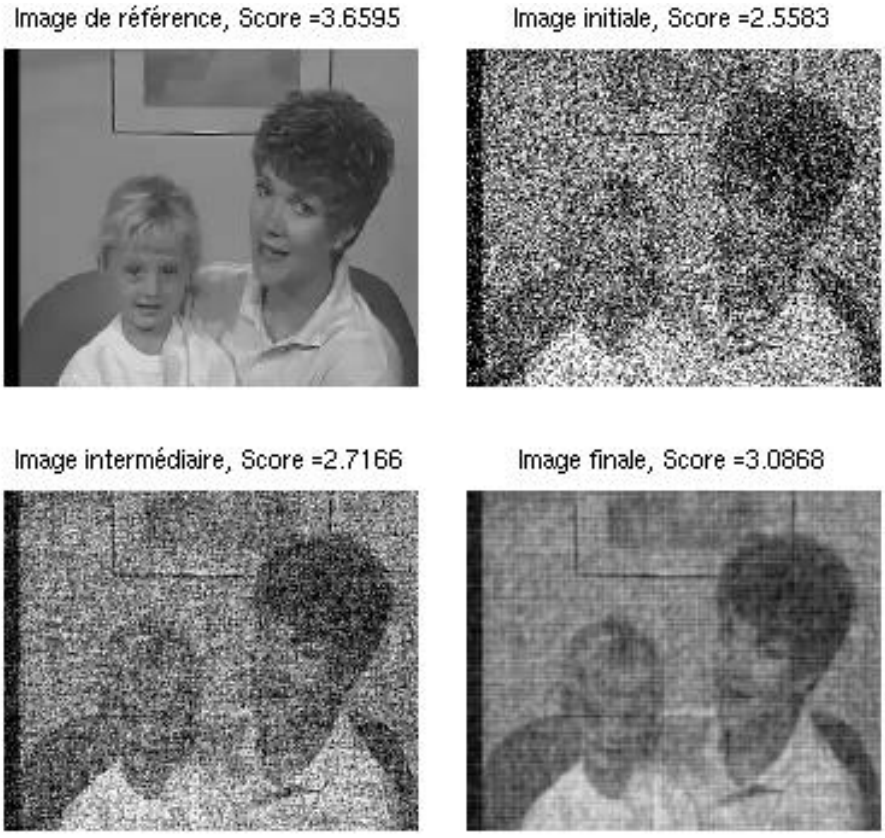


Figure 6.20 Résultats de dissimulation : scores estimés par l'OMQV pris comme fonction coût.

6.4 Mesure et test de la qualité d'une image

La technique de classification (chapitre 2) sera utilisée pour le test et la validation de la qualité visuelle de l'image décodée. Pour la validation des résultats expérimentaux liés au jugement sur la satisfaction du niveau de qualité des vidéos, l'approche sera similaire à celle développée au chapitre 4.

Les métriques traditionnelles telles que le PSNR sont plus utilisées dans le calcul métrique de la qualité objective. Cependant de telles métriques reflètent la différence absolue entre deux séquences et ignore la capacité du cerveau humain de compensation face à la dégradation de la qualité de l'image.

Les experts du Video Quality Expert Group (VQEG) ([13] et [28]) ont créé une spécification pour les tests de qualité subjective de l'image, ils l'ont soumise sous la recommandation ITU-R BT.500. Cette recommandation décrit la méthode de l'analyse subjective de la qualité d'images où des testeurs humains analysent des séquences d'images et leur donne une note qualitative.

Ces notes sont combinées, corrélées et reportées sous forme de score appelé Mean Opinion Score (MOS). C'est cette échelle de qualité qui a été utilisée dans les premiers chapitres du manuscrit et qui nous intéressera également dans ce chapitre vu que le jugement de validation ou rejet du niveau de qualité d'une image va être porté sur la note (MOS) qui lui a été attribué par le système de mesure de qualité implémenté au chapitre 4.

L'objectif ici est de prendre une décision pour une image donnée. Soit de valider l'image comme étant d'un niveau de qualité satisfaisant, soit plutôt de la rejeter. Pour cela il est important de fixer un ce seuil de qualité convenable.

6.4.1 Choix du seuil de qualité

Le seuil de qualité va permettre de trancher sur la validation ou le rejet de l'image une fois son niveau de qualité défini par l'outil de mesure de la qualité basé sur la régression décrit au chapitre 4. Il importe d'abord de se positionner sur une échelle d'évaluation appropriée.

Pour la validation des séquences d'images, le seuil défini est de 4 sur l'échelle continue de la qualité décrit dans le tableau 6.1.

Tableau 6.1 Echelle à 5 niveaux de qualité visuelle des images.

Qualité	Note	Dépréciation
Excellent	5	Imperceptible
Bon	4	Perceptible mais pas Ennuyeux
Moyenne	3	Un Peu Ennuyeux
Mauvais	2	Génant
Très mauvais	1	Très génant

Les valeurs heuristiques et nominales du MOS s'échelonnent de 1 à 5 :

- 4.4-5.0 – Très satisfait
- 4.0-4.3 – Satisfait
- 3.0-3.9 – Quelques utilisateurs satisfaits
- 2.0-2.9 – Beaucoup d'utilisateurs insatisfaits
- 1.0-1.9 – La plus part des utilisateurs insatisfaits

Ces valeurs heuristiques montrent que pour une image donnée, à partir d'un MOS supérieur ou égal à 4,3 le nombre d'utilisateurs ou d'opinions se déclarant satisfaits par la qualité de cette image est supérieur ou égale à la moyenne des utilisateurs. Les images ayant un MOS $\geq$ 4,3 seront donc validées directement. Par ailleurs à partir d'un MOS inférieur ou égal à 3,9 le nombre d'utilisateurs se déclarant satisfaits par la qualité de cette image est inférieur ou égale à la moyenne des utilisateurs. Les images ayant un MOS $\leq$ 3,9 seront donc considérées comme invalidées.

Nous fixons donc notre seuil s de qualité de l'image égale à 4 pris comme partie entière de la moyenne des deux valeurs ci-dessus. Une image sera donc acceptée par le système de contrôle si sa note sur 5 est supérieure ou égale à 4 et réinjectée dans l'étape de correction si sa valeur est strictement inférieure à 4.

6.4.2 Performances du système de validation des images

Cette section effectue une analyse des taux de réjection et d'acceptation du système de validation des images utilisant le seuil défini ci-haut.

Reprenons la base de données utilisée dans le chapitre 4 et analysons le taux de réjection et d'acceptation que le système de validation utilisant le seuil s défini ci-haut donnerait pour valider ses performances. La base de données a mesurée les scores de 62 images différentes déduites de 6 images originales au format CIF. Nous rappelons également que les résultats de notes de qualité obtenues par régression non linéaire étaient correctement corrélées aux notes données par des humains avec un coefficient de corrélation SCC = 0.929.

Lorsque l'outil de mesure de la qualité d'une vidéo (OMQV) basée sur la régression évalue la qualité d'une image en estimant son score S, le système de validation décide si l'image sera validé (si S $\geq$ 4, correspondant aux images représentées en couleur vert sur la Figure 6.15) ou si l'image est invalide (si S $\leq$ 4, correspondant aux images représentées en couleur rouge sur la Figure 6.15) dans quel cas elle repasse par la boucle de correction d'erreur comme indiquée dans la Figure 6.12.

Les performances du système de validation du décodage des images ainsi défini ont été évaluées par le calcul de deux métriques : le yield loss (YL) et le deflect level (D).

Calcul de l'élection de *yield loss* (YL):

Le paramètre noté YL appelé perte de rendement peut se définir comme le taux d'échec du système de validation sur les images ayant un niveau de qualité jugé comme satisfaisant. Sa formule est donnée par l'équation suivante :

$$Y_L = \frac{Nombre\ de\ vidéos\ de\ qualité\ satisfaisante\ mais\ ayant\ été\ invalidées\ par\ le\ test}{Nombre\ de\ vidéos\ de\ qualité\ satisfaisante} \quad (6.9)$$

Le YL trouvé pour 54 images utilisées lors du test est : YL = 0,083. Donc seulement 8% des images de qualité satisfaisante pour l'œil humain ont été considérées par le test comme étant de qualité Moyenne. Cette statistique montre que le système de test des images est assez fiable.

Calcul du deflect level (D):

Le paramètre noté D (deflect level) pouvant désigner une déviation de classement (classification manquée) peut se définir comme le taux d'échec du système de validation sur les images ayant un niveau de qualité jugé comme Moyenne par l'humain. Sa formule est donnée par l'équation suivante:

$$D = \frac{Nombre\ de\ vidéo\ de\ qualité\ insatisfaisante\ mais\ ayant\ été\ validées\ par\ le\ test}{Nombre\ de\ vidéo\ qui\ ont\ été\ validées\ par\ le\ test} \quad (6.10)$$

Le deflect level trouvé pour 54 images utilisées lors du test est : $\boldsymbol{D = 0.1419}$. Donc 14% des images qui ont été validées par le test seraient jugées par l'œil humain comme ayant une qualité satisfaisante. Ce chiffre reste assez petit pour confirmer que le système a de bonnes performances sur la validation des images.

6.5 Discussions et Perspectives

6.5.1 Conception d'une boucle de correction des erreurs dans le décodage MPEG

Dans ce paragraphe rédigé en guise de perspectives à la thèse, nous proposons un concept pour le contrôle de la qualité d'un décodage vidéo numérique. La vidéo est prise en entrée et traitée image par image. On reconsidère ici l'aspect temporel de la vidéo et les artefacts visuels de suroscillations et d'effets de bloc en plus du bruit et du flou. Le système intègre différents étage constituant les parties d'analyse et paramétrage des images, de priorisation des artefacts, de correction d'erreurs et de mesure et test du niveau de qualité des vidéos. La régression non linéaire (voir chapitre 4) ayant montré des résultats expérimentaux très intéressant sera retenue quant à la méthode de mesure de la qualité d'une vidéo.

Le contrôle de la qualité du décodage intègrerait une première étape de détection et priorisation des erreurs susceptibles de nuire à la qualité visuelle de la vidéo en entrée du décodeur. Cette étape considère 4 types d'artefacts visuels détectables : les effets de bloc, le flou, le bruit et les suroscillations. En fonction de l'artefact classé prioritaire, la deuxième étape de correction réalisée par le bloc "Error Concealment" se charge de corriger l'erreur. Cette deuxième étape est représentée dans la Figure 6.22 par le bloc "Eliminer l'artefact". Une fois la correction effectuée, une dernière étape va consister à mesurer le niveau de qualité de la trame décodée et le comparé à un seuil bien défini.

Le bloc "Error Concealment" réalise la dissimulation en appliquant l'algorithme de correction approprié suivant la nature de l'artefact prioritaire MAD. En effet, une fois l'artefact prioritaire MAD identifiée, une méthode de correction est appliquée suivant 2 choix existants : Si le MAD est le flou ou le bruit, une déconvolution (ou une convolution) avec une fonction gaussienne est appliquée à l'image ; S'il s'agit plutôt d'effets de bloc ou de suroscillations, la méthode spatio-temporelle est utilisée (voir Figure 6.12).

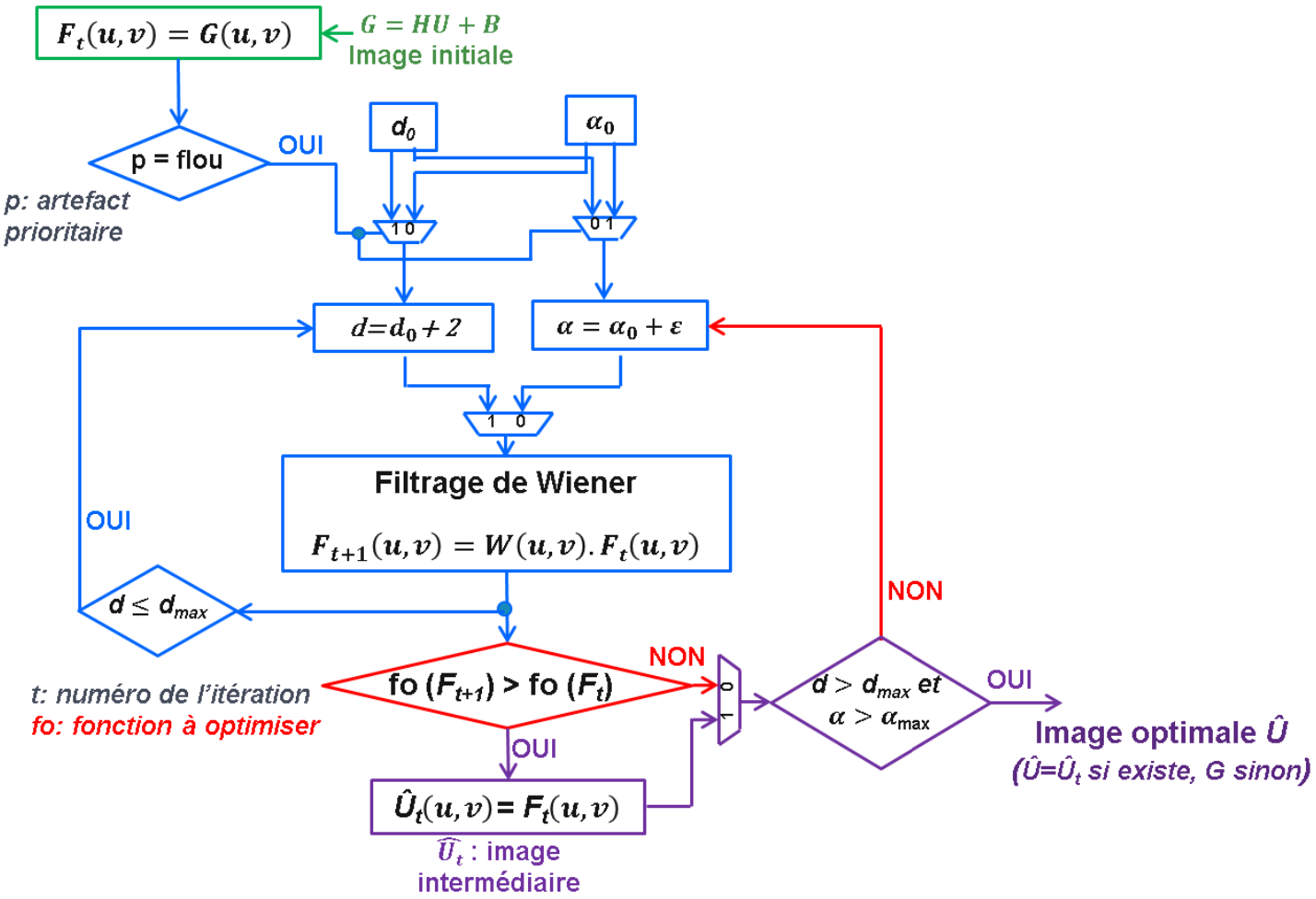


Figure 6.21 Schéma de la boucle de contrôle de la qualité du décodage.

6.5.2 Comparaison avec un outil commercialisé

L'entreprise Tektronix [1] a développé un outil de dernière génération pour l'analyse de la qualité des images dénommé PQA. Cet outil est basé sur les concepts du système visuel humain. Sa version PQA600 offre une suite de mesures objectives de qualité reproductibles qui correspondent étroitement à l'évaluation visuelle humaine subjective. Ces mesures fournissent des renseignements précieux pour les ingénieurs travaillant pour optimiser la compression et le décodage numérique de vidéos ou d'images.

Sur la technique utilisée pour l'évaluation de la qualité des vidéos ou d'images, le PQA se base également sur le SVH, tout comme l'OMQV. Les deux techniques d'évaluation de la qualité des images utilisent des résultats de la statistique du réseau comme le taux de perte de paquets. Toutefois, la pertinence des résultats de l'OMQV est montrée dans l'élaboration de ses performances (ch. 2, 3, et 4) par comparaison directe avec l'évaluation subjective de la qualité des vidéos effectuée par des humains. Les deux expériences travaillent sur la base du décodage numérique de la famille MPEG.

Sur la correction des erreurs, aucune démarche expérimentale n'a été apportée (du moins au public) pour montrer comment le PQA aide à la compression et le décodage numérique de vidéos. En revanche, ce manuscrit a démontré (paragraphe 6.3.2) que l'algorithme de

correction d'erreurs conçu ici utilise en temps réel l'outil de mesure de la qualité des vidéos (OMQV) pour s'assurer que la qualité recherché dans l'optimisation de l'image soit bien appréciée par l'être humain. Cela constitue le principal avantage de l'algorithme de correction proposé.

6.6 Conclusion

Dans ce chapitre, une contribution à été apportée sur l'évaluation de la qualité d'une image par la création d'une évaluation objective de la qualité vidéo en corrélation avec le système visuel humain. Une démonstration a été faite pour montrer comment nous assurer une bonne qualité visuelle aux images décodées grâce à l'algorithme de correction d'erreur assisté par l'OMQV basé sur les RNAs. Il faut noter une fois de plus que la méthode élaborée ici est applicable à tout type de format d'image.

Une boucle de contrôle de la qualité d'images intégrée dans un décodage numérique d'images a été élaborée. L'utilisation de la régression non linéaire a permis d'incorporer dans la boucle de décodage numérique d'images un outil de mesure de la qualité d'une image capable d'assurer une évaluation numérique de la qualité d'une image quasiment similaire en termes de performances au jugement humain sur la qualité des images. Une étude d'un seuil de qualité universellement reconnu sur la base d'expériences subjectives a permis d'établir un test de validation des images décodées de façon à satisfaire à la perception visuelle humaine sur la qualité des images décodées. Trois techniques de dissimulation d'artefacts ont été proposées selon le type d'artefact considéré comme prioritaire vis-à-vis de l'impact sur détérioration de la qualité de l'image.

Conclusion et Perspectives

Dans cette thèse, des apports ont été apportées dans le domaine du décodage vidéo numérique, notamment sur l'évaluation objective de la qualité visuelle d'une vidéo ou d'une image et la dissimulation des artefacts visuels à la sortie des images décodées.

Les trois techniques développées pour la mesure de la qualité visuelle d'une image ou d'une vidéo constituent un apport essentiel de cette thèse. Les méthodes d'évaluation de la qualité, autrefois essentiellement subjectives, consistaient à soumettre une vidéo (ou une image) sous l'appréciation d'un ensemble d'individus qui affectaient une note à cette vidéo (ou à cette image) suivant son niveau de qualité visuelle. Un score de qualité de la vidéo (ou de l'image) noté MOS était ensuite calculé comme moyenne des différentes notes données par les observateurs. Cette méthode a montré des limites de point de vue pratique parce qu'elle n'est pas temps-réel et est trop couteuse. La recherche s'est ensuite orientée vers des méthodes objectives. Nous y avons contribué par les trois propositions apportées.

La première technique basée sur la classification s'inspire d'une observation sur les expériences subjective. En effet on remarque que lors des expériences subjectives les individus évaluent les images en les regroupant par niveaux de perception visuelle des artefacts sur l'image. On retrouve ainsi cinq principales classes de qualité : très gênant, gênant, un peu ennuyeux, perceptible mais moins ennuyeux et imperceptibles. Cinq classes de qualité ont été définies, à partir de l'échelle de qualité à 5 niveaux du groupe ITU (Très mauvais, Mauvais, Moyenne, Bon, Excellent) correspondant respectivement à ces cinq niveaux de perception visuelle des artefacts. On obtient un très bon classement, avec un taux de réussite moyen de 77,95%.

La deuxième technique basée sur les réseaux de neurones artificiels (RNAs) s'inspire tout simplement du fonctionnement des réseaux de neurones du cerveau humain. La méthode utilise un apprentissage supervisé du RNA en prenant en entrée les paramètres de qualité de l'image PMR, PSNR, SI et blurMetric et génère en sortie un score. La corrélation obtenue entre les scores générés et les MOS est satisfaisante, avec un coefficient Spearman de 0.97. Cette technique a montré une efficacité directement comparable aux MOS résultants du jugement de l'intelligence humaine.

La troisième technique d'EQV élaborée utilise la régression non linéaire (RNL), un outil avancé de l'analyse statistique. Le coefficient de corrélation obtenu avec les MOS est de $R^2 = 0.9406$. Ce résultat démontre des résultats satisfaisants. La RNL nous laisse la possibilité d'affiner l'estimation des scores de qualité en recherchant des modèles de régression plus

représentatifs de la corrélation de la qualité visuelle des images avec chacune des métriques utilisées.

L'utilisation de l'OMQV a permis d'une part d'incorporer dans la boucle de décodage numérique un système capable d'assurer une évaluation numérique de la qualité d'une image quasiment similaire en termes corrélation vis-à-vis du jugement humain sur la qualité des images. D'autres parts, on a montré qu'elle permet d'augmenter les performances d'un algorithme de correction d'erreurs de décodage dans une image. Une étude d'un seuil de qualité universellement reconnu sur la base d'expériences subjectives a permis d'établir un test de validation des images décodées de façon à satisfaire à la perception visuelle humaine sur la qualité des images décodées. Trois techniques de dissimulation d'artefacts ont été proposées selon le type d'artefact considéré comme prioritaire vis-à-vis de l'impact sur détérioration de la qualité de l'image.

Un concept de boucle de contrôle de la qualité visuelle d'une image a été étudié. Dans une première partie de cette boucle de contrôle, la qualité visuelle de la vidéo ou de l'image est mesurée par des techniques basées sur l'analyse statistique. Dans une seconde partie, un seuil de qualité est évalué à partir du quel les vidéos ou les vidéos de qualité insatisfaisantes peuvent être soumises à un bloc de dissimulation des artefacts.

Ce projet a ouvert des horizons à des travaux futurs dont la validation de l'OMQV par une autre base de données telle que CSIQ ; l'implémentation matérielle du contrôle de la qualité des images dans un décodeur en prenant en considération d'autres types d'artefact (effets de bloc, suroscillations) ; le rajout de la métrique TI (temporal Index) dans les paramètres à considérer dans la conception d'un outil d'évaluation des vidéos qui pourra prendre en compte les aspects temporelles afin de refléter mieux cette dimension non négligeable dans le jugement du niveau de qualité d'une vidéo et enfin l'implémentation matérielle de l'OMQV en prenant en considération d'autres attributs (paramètres vidéo) tels que la texture.

Bibliographie

[1] (s.d.). Récupéré sur Tektronix website : http://www.tek.com/vidéo-quality-monitors

[2] (2008). Subjective Vidéo Quality Assessment Methods for Multimedia Applications. Dans *ITUT Recommendation* (p. 910).

[3] Avcibas, I., Sankur, B., & Sayood, K. (2002). "Statistical Evaluation of Image Quality Measures". *Journal of Electronic Imaging, vol. 11*, 20-23.

[4] Beghdadi, A., & Deriche, M. (2000). "Features extraction from fingerprints using frequency analysis". *IEEE Workshop On Signal Processing and its Applications.*

[5] Belfiore, S., Grangetto, M., Magli, E., & Olmo, G. (2003). Spatio-temporal video error concealment with perceptually optimized mode selection,". *Int. Conf. on Multimedia and Expo*, *vol. 3*, pp. 6-9.

[6] Bovik, A. C., & Liu, S. (2001). "DCT-domain blind measurement of blocking artifacts in DCT-coded images,". *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing.*

[7] Chandler, D. M., & Hemani, S. S. (2007). online supplement to ‘vsnr: A visual signal-to-noise ratio for natural images based on near-threshold and suprathreshold vision.

[8] Charrier, C., Lézoray, O., & Lebrun, G. (2009). Mesure de la qualité d’images couleur par combinaison de classifieurs. *Colloques sur le traitement du Signal et des Images.* GRETSI.

[9] Chetouani, A., Beghdadi, A., Chen, S., & Mostafaoui, G. (2010). A novem free reference image quality metric using neural network approach. *Proc. Int. Workshop Video Process. Qual. Metrics Cons. Electrn*, (pp. 1-4).

[10] Chetouani, A., Mostafaoui, G., & Beghdadi, A. (2009). "A new Free Reference Image Quality Index Based on perceptual Blur Estimation,". *IEEE Pacific-Rim Conference on Multimedia.*

[11] Chetouani, A., Mostafaoui, G., & Beghdadi, A. (2009). “Deblocking method using a perceptual recursive filter,". *IEEE International Conference on Image Processing.*

[12] Corchs, S., Gasparini, F., & Schettini, R. (2014). Noisy images-JPEG compressed: subjective and objective image quality evaluation. *Image Quality and System Performance XI. 9016, 90160V-1.* Milan: SPIE-IS&T Electronic Imaging.

[13] Corriveau, P., & Webster, A. (2003). *Final report from the video quality experts group on the validation of objective models of video quality assessment, phase II.* Video Quality Expert Group.

[14] Coudoux, F. X., Gazalet, M. G., & Corlay, P. (1998). "Reduction of blocking effet in DCT-coded images based on a visual perception criterion,". *Signal Processing: Image Communication*, (pp. 179-186).

[15] Crete, F., Dolmiere, T., Ladret, P., & Nicolas, M. (2007). The Blur Effect: Perception and Estimation with a New No-Reference. *SPIE Electronic Imaging Symposium Conf Human Vision and Electronic Imaging.* San Jose: SPIE.

[16] De Simone, F., Goldmann, L., Baroncini, V., & Ebrahimi, T. (2009). Subjective evaluation of JPEG XR image compression. *SPIE Optics and Photonics. 7443.* San Diego: Proceedings of SPIE.

[17] De Simone, F., Naccari, M., Tagliasacchi, M., Dufaux, F., Tubaro, S., & Ebrahimi, T. (2009). Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel. *Proc. Int. Conf. QoMEX.*

[18] De Simone, F., Tagliasacchi, M., Naccari, S., Tubaro, S., & Ebrahimi, T. (2010). A H.264/AVC video database for the evaluation of quality metrics. *Proc. of IEEE Inter. Conf. on Acoustic Speech and Signal Processing (ICASSP).* Dallas.

[19] Eden, A. (2008, Sept.). No-Reference Image Quality Analysis for Compressed Video Sequences. *IEEE transaction on broadcasting, 54*(N°3).

[20] Egiazarian, K., Astola, J., Ponomarenko, N., Lukin, V., Battisti, F., & Carli, M. (2006). Two new full-reference quality metrics based on HVS. *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, *CD-ROM*, p. 4. Scottsdale.

[21] Elad, M. (2002). On the origin of the bilateral filter and ways to improve. *IEEE Trans. Image Process*, 1141-1151.

[22] Farias, M. C. (2004). *"No-Reference and Reduced Reference Video Quality Metrics: New Contributions,."* Ph.D Dissertation, University of california, Dept. of Electrical and Computer Engineering, Santa barbara.

[23] Farias, M. C., Foley, J. M., & Mitra, S. K. (2007). "Detectability and Annoyance of Synthetic Blocky, Blurry, Noisy, and Ringing Artifacts,". *IEEE trans. Int. Conf. on Signal Processing*, (pp. 2954-2964).

[24] Farias, M. C., Moore, M. S., Foley, J. M., & Mitra, S. K. (2005). "Detectability and Annoyance of Synthetic Blocky and Blurry Video Artifacts". *Accoustic, Speech, and Signal Processing IEEE Int. Conf. 2*, pp. 553-556. IEEE.

[25] Ferzli, R., & Karam, J. L. (2009). "A No-Reference Objective Image Sharpness Metric Based on the Notion of just Noticeable Blur,"., *vol. 18*, pp. 717-728.

[26] Friebe, M., & Kaup, A. (2006). "3D-deblocking for Error Concealment in Block-based Video decoding Systems. *Proc. on Picture Coding Symposium.*

[27] Geisler, W. (2007). "Visual perception and the statistical properties of natural scenes" . *ANN Rev. Neuroscience*.

[28] Gelasca, E. D. (2005). *Full-reference objective quality metrics for video watermarking, video segmentation and 3D model watermarking.* Luisane: EPFL.

[29] Group, V. Q. (2000). *Final report from the video quality experts group on the validation of objective models of video quality assessment.* VQEG.

[30] Hsu, V. N., & Lin, C. -J. (2002). "A comparison of methods for multiclass support vector machines,". *IEEE Trans. on Neural Networks*, *13*, pp. 415-425.

[31] ITU-R. (1998). *"Methodology for the subjective assessment of the quality of television pictures"*.

[32] ITU-R Recommendation BT.500-11. (2002). *"Méthodologie d'évaluation subjective de la qualité des images de télévision,."* Génève: UIT.

[33] Jang, I. H., Kim, N. C., & So, H. J. (2003). "Iterative blocking artifact reduction using a minimum mean square error filters in wavelet domain". *Signal Processing*, (pp. 2607-2619).

[34] Kandel, Shwartz, & Jessel. (2005). Introduction to artificial neural networks. Dans Paplinski, *Principles of Neural Science.*

[35] Khereddine, R., Simeu, E., & Mir, S. (2007). "Utilisation des Techniques de régression pour le test et le diagnostic des composantes RF". *Journées GDR SoC-SiP.* Paris.

[36] Kupka, L., Simeu, E., Stratigopoulos, H., Rufer, L., Mir, S., & Tumova, O. (2008). Signature analysis for MEMS pseudorandom testing using neural networks. *12th IMEKO Joint Symposium on Man Science and Measurement*, (pp. 321-325). Annecy.

[37] Li, Q., & Wang, Z. (2009). “Reduced-Reference Image Quality Assessment Using Divisive Normalization-Based Image Representation”. *IEEE J. Selected Topics in Signal Proc., 3*(N° 2), 202-211.

[38] Li, X., & al. (2002). "Blind image quality assessment. *Proc. IEEE Int. Conf. Image proc.*, *1*, pp. 449-452.

[39] Liu, H., & Heynderickx, I. (2009). "A Perceptually Relevant No-Reference Blockiness Metric Based on local Image Characteristics,". *EURASIP Journal on Advances in Signal Processing*.

[40] Liu, H., Klomp, N., & Heynderickx, I. (2010). "A No-Reference Metric for Perceived Ringing Artifacts in Images,". *IEEE Trans. on Circuits and Systems for Video Technology*, 529-539.

[41] Liu, J., & Liang, D. (2005). A Survey of FPGA-based hardware implementation of ANNs. *Proc. Int. Neural Networks Brain*, *2*, pp. 915-918.

[42] Marziliano, P., Dufaux, F., Winkler, S., & Ebrahimi, E. (2002). "A no-reference perceptual blur metric," . *IEEE International Conference on Image Processing*, (pp. 57-60).

[43] Marziliano, P., Dufaux, F., Winkler, S., & Ebrahimi, T. (2004). "Perceptuel blur and ringing metrics: application to JPEG2000,". *Signal Processing: Image Communication*, *vol. 19*, pp. 163-172.

[44] Massachussetts Institute of technology. (2004). *Image processing toolbox documentation.* Récupéré sur The MathWorks, Inc: www.mathworks.fr/products/demos/image/ipexblind/ipexblind.html#2

[45] Moorthy, A. K., & Bovik, A. C. (2010). "A two-step framework for constructing blind image quality indeces". *IEEE Signal Processing Letters*, (pp. 587-599).

[46] Moorthy, A. K., & Bovik, A. C. (2011, Dec.). “Blind image quality assessment: From natural scene statistics to perceptual quality”. *IEEE Trans. Image Process., 20*(N° 12), 3350-3364.

[47] Muntean, G. M., Perry, P., & Murphy, L. (2005). Subjective assessment of the quality-oriented adaptive scheme. *IEEE Trans. Broadcast., 51*(N° 3), 276-286.

[48] Paplinski, A. P. (2005). “Concept neurons – Introduction to artificial neural networks”. Dans Kandel, Schwartz, & Jessel, *Princoples of Neural Sciences.*

[49] Photoshop. (s.d.). *Photoshop tips & tricks*. Consulté le 08 7, 2012, sur graphic design & publishing center: http://www.graphic-design.com/photoshop/photoshop-noise-reduction

[50] Ponomarenko, N., Lukin, V., Egiazarian, K., Astola, J., Carli, M., & Battisti, F. (2008). Color Image Database for Evaluation of Image Quality Metrics. *Proc. of the Inter. Workshop on Multimedia Signal Processing*, (pp. 403-408).

[51] Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., & Battisti, F. (2009). "TID2008 - A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics". *Advances of Modern Radioelectronics*, *Vol. 10*, pp. 30-45.

[52] Reibman, A. R. (2009, Sept.). "Monitoring the video quality inside a network". *AT&T Labs -Research Florham park*. Santa Clara, NJ.

[53] Sambhare, A. R. (2003). *"Detecting Artifacts and Textures in Wavelet Coded Images."* ECE.

[54] Sheikh, H. R., Bovik, A. C., & Cormack, L. K. (2005). "No-Reference Quality Assessment Using natural Scene Statistics: JPEG2000". *IEEE transactions on Image Processing, vol. 14*.

[55] Sheikh, H. R., Sabir, M. F., & Bovik, A. C. (2006). “A statistical evaluation of recent full reference image quality assessment algorithms”. *IEEE Transactions on Image Processing, vol. 5*(N° 11), 3441-3452.

[56] Sheikh, H. R., Wang, Z., Cormack, L., & Bovik, A. C. (s.d.). *Live image quality assessment database release 2*. Récupéré sur http://live.ece.utexas.edu/research/quality

[57] Simeu, E. (2005). ‘’Test et Surveillance Intégrés des Systèmes Embarqués,’’ *HDR, Grenoble: Université Joseph Fourier, pp. 69-70.*

[58] Simonceli, E. P., Freeman, W. T., Adelson, E. H., & Heeger, D. J. (1992). “Shiftable multiscale transforms”. *IEEE Transactions on Information Theory, 38*(N° 2), 587-607.

[59] Stratigopoulos, H., & Makris, Y. (2008). Error moderation in low-cost machine-learning-based analog/RF testing. *IEEE Trans. on Computer-Aided Design of Integrated Circuirts and Systems, 27*(2), 339-351.

[60] Su, L., Zhang, Y., Gao, W., Huang, Q., & Lu, Y. (2004). Improved Error Concealment Algorithms Based on H.264/AVC Non-nrmative Decoder. *Proc. Int. Conf. Multimedia Expo (ICME)*, (p. 1671).

[61] Tang, H., Shi, H., & Zhao, H. (2008). An Improved Error Concealment Algorithm for Intra-frames of H.264/AVC. *Image and Signal Processing. 4*, pp. 485-488. IEEE.

[62] Tomasi, C., & Manduchi, R. (1998). Bilateral filtering for gray and color images. *Proceedings of the ICCV*, (pp. 836–846.).

[63] Tong, H., Li, M., Zhang, H., & Zhang, C. (2004). "Blur detection for digital images using wavelet transform,". *IEEE International Conference on Multimedia and Exposition*, *vol. 1*, pp. 17-20.

[64] Vlachos, T. (2000, June 22nd). “Detection of blocking artifacts in compressed video”. *Electronics letters, 36*(N° 13).

[65] Wang, Z., & Bovik, A. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 81-84.

[66] Wang, Z., Bovik, A. C., & Evans, B. L. (2000). Blind measurement of blocking artifacts in images. *IEEE Int. Conf. Image Processing*, *3*, pp. 981-984.

Annexes

Annexe 1 : Liste des publications

Conférences internationales

1 B. Ekobo Akoa, E. Simeu, F. Lebowsky, "Using statistical analysis and artificial intelligence tools for automatic assessment of video sequences". Proc. SPIE 9015, Color Imaging XIX: Displaying, Processing, Hardcopy, and Applications, San Francisco, Californie, Etats-Unis, 8/01/2014

2 B. Ekobo Akoa, E. Simeu, F. Lebowsky, "Using Classification for Video Quality Evaluation". 25th IEEE International Conference on Microelectronics (ICM'13), Beyrouth, Lyban, 15-18/12/2013

3 B. Ekobo Akoa, E. Simeu, F. Lebowsky, "Video decoder monitoring using non-linear régression". 19th IEEE International On-Line Testing Symposium (IOLTS'13), La Canée, Crête, 8-10/07/2013

4 B. Ekobo Akoa, E. Simeu, F. Lebowsky, "Using Artificial Neural Network for Automatic Assessment of Video Sequences".27th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA'13), Barcelone, Espagne, 25-28/03/2013

Conférence nationale

5 B. Ekobo Akoa, E. Simeu, F. Lebowsky, "Utilisation des Techniques Avancées de l'Analyse Statistique dans le Décodage Vidéo Numérique". Journées scientifiques Semba'2013, Saint Germain au Mont d'Or, France, 4-5/04/2013.

Annexe 2: Détection et classification d'artefacts de compression vidéo

A.1 Introduction

Cette annexe est le résumé du rapport de stage de Mourad GOUMRHAR, durant l'été 2012 au sein de l'équipe Algo dans la division DCG, et plus précisément au sein de *Connected Home Division (CHD)* pour laquelle cette thèse a été réalisée. Ce travail contribue au diagnostic des artefacts visuels établi dans les quatre premiers chapitres de la thèse.

A.1.1 Problématique

Le produit phare de CHD est la Set-Top Box, ou décodeur TV. Ce type d'appareils est de plus en plus répandu sur le marché comme moyen de réception des chaines de télévision et des vidéos numériques, utilisé conjointement avec une Box ADSL.

La Set-top Box assure notamment le décodage des flux vidéo qui sont compressés avant leur transmission afin de réduire le volume des données. La plupart des techniques de compression de vidéos sont avec pertes, dans le sens où la qualité de la vidéo est volontairement dégradée sans possibilité de retrouver la qualité d'origine. En l'occurrence, les codecs (ou codeur/décodeurs) concernés sont H.264/MPEG-4 AVC et le futur H.265 aussi appelé HEVC (*High Efficiency Video Coding*) en cours de standardisation.

Les algorithmes de compression avec perte utilisés introduisent souvent des défauts visuels qui sont inhérents aux opérations effectuées, indépendamment du contenu de la vidéo qui peut présenter intrinsèquement certaines de ces distorsions et qui proviennent, par exemple, de la réutilisation de contenu déjà compressé. Dans un souci de qualité finale de l'image, la plupart des décodeurs utilisent des filtres de post-traitement visant à pallier certains de ces artefacts de manière spécifique.

Les modules de corrections sont configurés une fois pour toutes à la fabrication selon la commande du client sur le niveau Low, Medium ou High ou sont tout simplement désactivés. Cependant, le comportement des filtres dépend grandement du contenu des images. Une telle configuration statique peut donc introduire des artefacts en essayant de dissimuler certains défauts ; par exemple, un filtre de réduction des blocs visibles sur les images compressées en JPEG va augmenter le flou.

C'est donc dans cette optique qu'il y a tout intérêt à avoir une solution de décodage et de correction paramétrables qui s'adapteraient au contenu en examinant la qualité des images.

L'idée proposée est alors d'introduire, comme schématisé en Figure A.1 un détecteur d'artefacts dont le but est d'identifier, parmi les distorsions spatiales que l'on cherche à corriger, celles qui sont présentes dans les séquences d'images à visualiser et de détecter lesquelles sont les plus visibles et qui seraient donc à traiter en priorité.

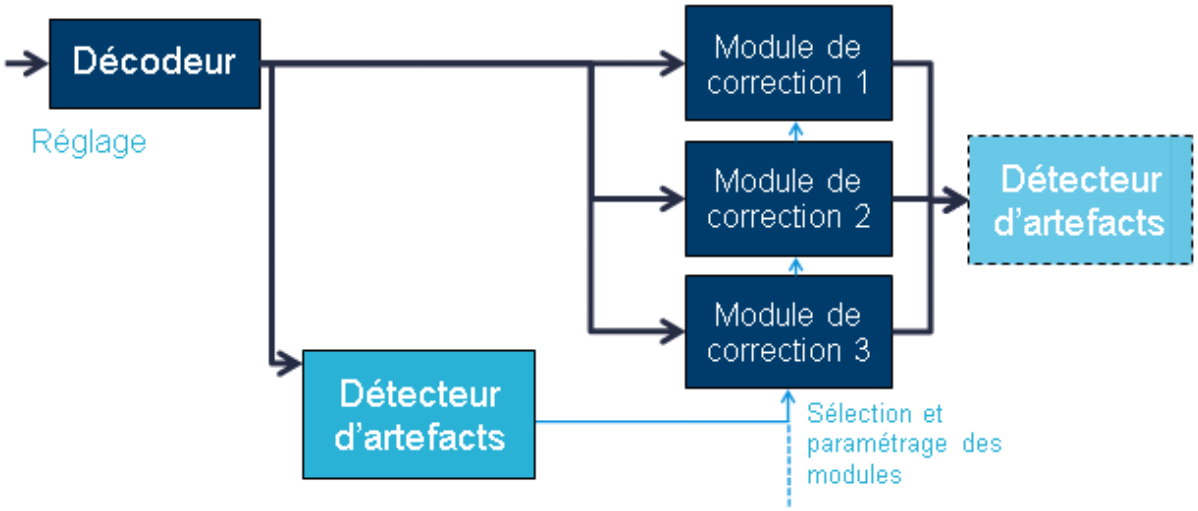


Figure A.0.1 Positionnement et fonctions du détecteur d'artefacts.

Une fois l'artefact identifié, il serait possible d'activer le module de correction spécifique de la distorsion considérée et de fixer sa finesse. De plus, la qualité en sortie du décodeur pourrait être évaluée en temps réel afin de régler de manière plus précise ses paramètres en mode retour de boucle.

Il serait tout aussi envisageable d'utiliser un détecteur en sortie des modules de correction afin de mesurer la qualité des images finales et de vérifier que les corrections potentielles n'ont pas introduit de défauts majeurs.

Le but du stage fut donc d'étudier la faisabilité de ce genre de détecteurs et leur performance. Afin de reconnaitre et faire distinction entre les différentes classes de distorsions, on se propose d'utiliser un classifieur de type *Support Vector Machine (SVM),* ou Machine à Vecteurs Support.

Dans cette annexe à la thèse sont d'abord explicitées les distorsions propices aux décodeurs de la famille MPEG ainsi que leur origine. Ensuite, la méthode de classification sera exposée, à savoir les SVM, ainsi que les contraintes qu'elle impose. Suivra une étude de ce qui existe en matière d'évaluation de qualité utilisant des *SVM* avant d'exposer et d'analyser la méthode développée pour le projet.

A.2 Analyse

A.2.1 Origine et nature des artefacts ciblés

Les artefacts à détecter sont les principales catégories de distorsions connus dans le décodage vidéo numérique. Les décodeurs H.264/MPEG-4 AVC (*Advanced Video Coding*) et son tout récent successeur H.265/HEVC (*High Efficiency Video Coding*) introduisent des distorsions dues à des procédés liés aux récents algorithmes de décodages.

Ce travail ne s'est intéressé qu'aux distorsions spatiales; négligeant ainsi toutes formes de distorsions temporelles.

a) **Compression spatiale des images**

Une vidéo étant constituée de plusieurs séquences (trames) d'images, le décodage vidéo numérique va d'abord consister en une compression de ces trames d'image comme l'indique la Figure A.2.

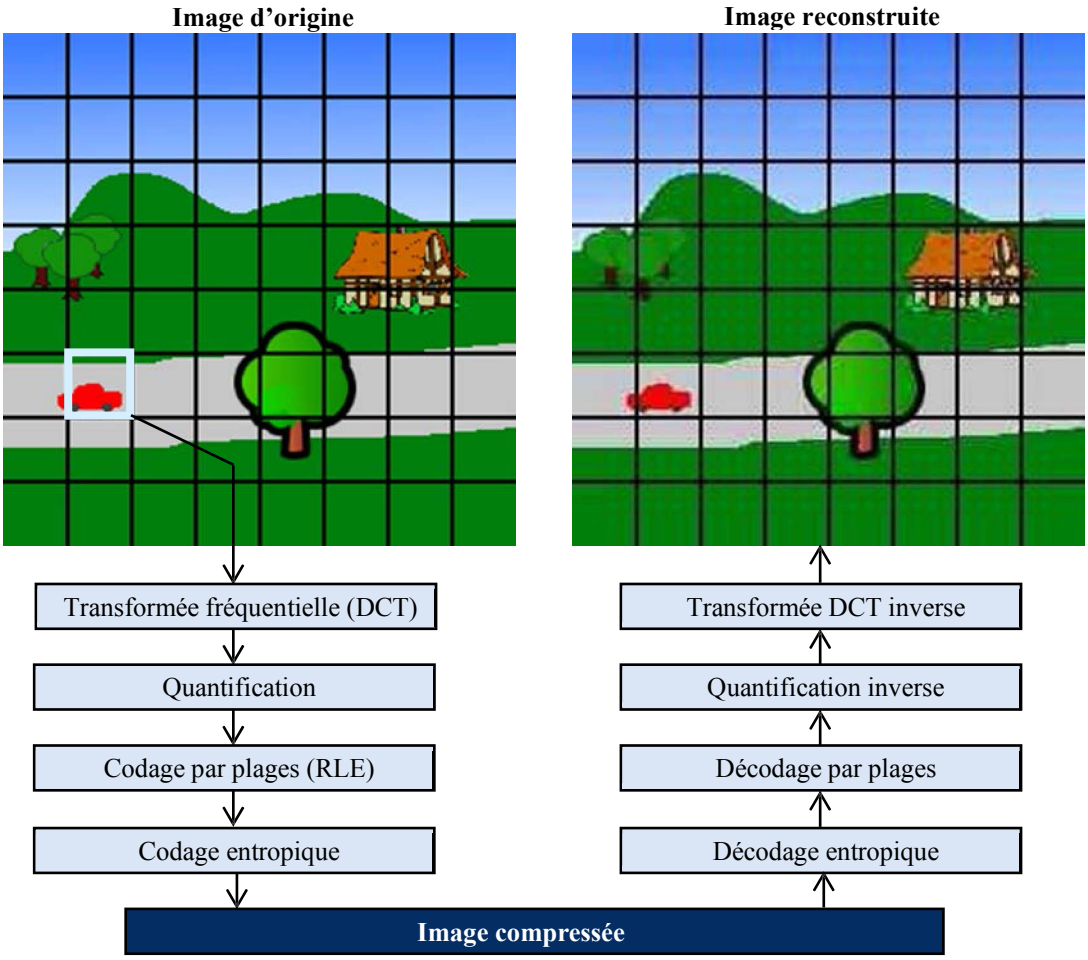


Figure A.0.2 Compression des images par le décodeur JPEG.

Le procédé de compression spatiale des images se déroule de la façon suivante :

- Décomposition de l'image en macroblocs des blocs 4x4 pixels ;
- Calcul de la transformée DCT (*Discrete Cosine Transform*) de chaque bloc. Permettant de séparer les informations perceptuelles (basse fréquences) des détails (haute fréquences), moins sensibles à l'œil humain;
- Quantification des coefficients DCT en haute précision (moins de bits utilisés) pour les basses fréquences et basse précision pour les hautes fréquences.

Les distorsions sont principalement causées par cette dernière étape, vu qu'on ne retrouve pas la précision exacte en haute fréquence. En effet, la quantification effectue une division des

coefficients DCT par une table dite de quantification, constituée des coefficients fixés par le CODEC (le couple encodeur/décodeur) selon la précision choisie, puis va arrondir à une valeur entière entraînant ainsi une perte d'informations.

- Le codage par plage va ensuite réordonnancer en zigzag (Figure A.3) les coefficients nuls consécutifs et code leur nombre au lieu de coder les valeurs nulles individuellement.

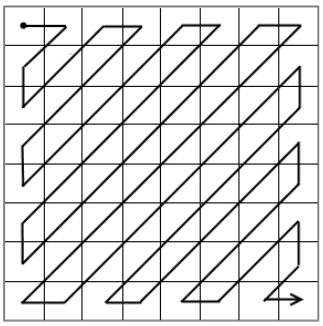


Figure A.0.3 Réordonnancement en zigzag.

- Le codage entropique est alors appliqué sur les symboles représentant les coefficients DCT quantifiés. Il consiste à remplacer les symboles les plus fréquents par des mot-codes de faible longueur, et les moins fréquents par des mot-codes plus longs.

b) **Les types d'artefacts ciblés**

Les artefacts concernés sont en plus des 4 distorsions mentionnées dans le manuscrit (le flou, le bruit, effets de bloc et suroscillations) les erreurs de transmission encore connus sous le nom anglais *Fast Fading*. Le travail décrit ici s'est toutefois limité aux artefacts flou et effets de bloc.

Pour des raisons de prix ou des besoins expérimentaux, certains CODEC sont conçu avec des contraintes limitant exclusivement un ou plusieurs types d'artefacts parmi ces principales catégories. C'est pour ça que le classifieur SVM a été choisi dans ce travail pour sa nature multi-classe et sa facilité d'implémentation par rapport par exemple aux RNA (cf. ch. 3).

A.2.2 Détection et identification par une SVM

Les Machines à Support de Vecteurs, SVM classées parmi les techniques d'apprentissage supervisé, analysent et reconnaissent des données ou des motifs en vue d'une classification (cf. ch. 2) ou de régression (cf. ch. 4).

a) **Principe de fonctionnement**

Les SVM fonctionnent selon 2 modèles :

- **Echantillons linéairement séparables** : établissement d'une frontière (ou hyperplan) entre les échantillons des différentes classes de façon à maximiser la marge (Figure A.4), entendue comme la distance aux vecteurs supports (VS), qui sont les échantillons les plus proches.

La généralisation, qui est une séparation à de nouveaux échantillons est alors effectuée par un calcul de leur position par rapport à cet hyperplan.

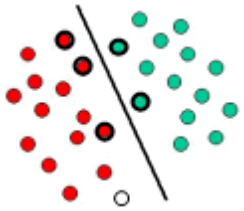


Figure A.0.4 SVM avec échantillons linéairement séparables : VS en gras et échantillons à classifier en blanc

- **Echantillons non linéairement séparables :** projection vers un espace vectoriel de plus grande dimension (voire de dimension infinie) en utilisant une transformation non-linéaire dite fonction noyau (Figure A.5). Une frontière de séparation est ensuite calculée et la généralisation se fait comme précédemment.

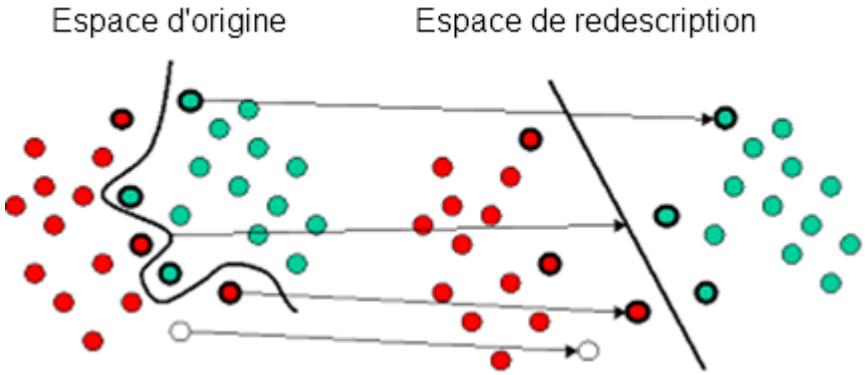


Figure A.0.5 SVM avec des échantillons non linéairement séparables.

Comme le montre la Figure A.6, la classification va estimer les classes à partir d'un vecteur représentatif à valeurs réelles préalablement déduit par calcul de la séparation de l'espace vectoriel en hyperplans.

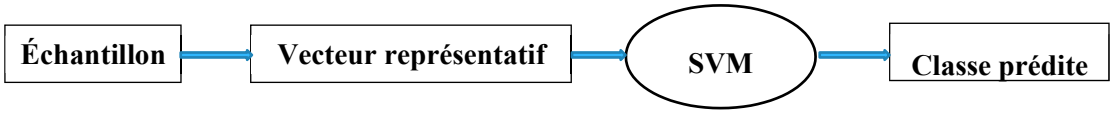


Figure A.0.6 Schéma de classification à l'aide des SVM.

b) **Apprentissage et Test d'une SVM**

L'apprentissage consiste pour une SVM à mémoriser des motifs en vu de leur reconnaissance alors que le test permet de valider le modèle de classification.

Durant l'apprentissage, la SVM prend en entrées des vecteurs et les classes qui les correspondent, ce qui permet de configurer la fonction noyau et ses paramètres. Les fonctions noyau φ disponibles sont de type linéaire, polynomial, fonctions gaussiennes à base radiale et fonctions sigmoïdes.

La fonction à base radiale qui est la plus utilisée se présente comme suit :

$$\phi(x_1, x_2, \dots, x_n) = e^{-\gamma|x_i - x_j|^2}$$

La constante γ fait partie des paramètres à configurer pour optimiser la classification, ainsi qu'un paramètre coût noté *C*, qui est essentiellement fonction des vecteurs échantillons.

Le pourcentage des échantillons bien classés par la SVM lors durant la phase d'apprentissage en mesure l'efficacité. Il correspond au nombre d'échantillons ayant été bien classé sur le nombre total d'échantillons réservés à l'apprentissage.

Durant la phase de test les vecteurs (échantillons) réservés au test ainsi que leurs classes respectives sont soumis la SVM. Ces vecteurs doivent être différents de ceux réservés à la phase d'apprentissage et l'ensemble doit constituer une distribution homogène par rapport aux classes d'appartenance.

La SVM donne alors en sortie une classe estimée pour chaque échantillon de test et la compare à la classe indiquée en entrée. Ce qui donne une erreur qui est la différence entre la classe réelle (fournie en entrée) et la classe estimée (restituée en sortie). Le test est valide en mesure que cette erreur est proche de zéro. De même qu'en apprentissage, un pourcentage d'efficacité va permettre de mesurer la capacité de généralisation du modèle.

c) **SVM multi-classes**

L'utilisation des SVM a l'avantage de traiter plus simplement les problèmes multi-entrées. Il suffit d'écrire chaque vecteur échantillon avec des paramètres représentatifs de chaque entrée. Pour n classes considérées, la classification peut alors se faire de deux façons :

- La méthode "un contre tous" où une classe est mise à part et les autres n-1 sont fusionnées en une même classe. Le modèle définitif sera celui qui génèrera la plus haute probabilité d'appartenance à une classe, pour un échantillon donné.
- La méthode "un contre un" où n(n-1)/2 modèles sont créés. La classe d'un échantillon est celle qui aura la plus grande probabilité d'appartenance de cet échantillon à cette classe.

d) **Application des SVM pour la détection d'artefacts**

La méthode utilisée ici appartient à la famille des techniques de détection et de mesure des distorsions dites sans référence (cf. ch.1, section 1.2). L'utilisation d'une SVM comme technique d'intelligence artificielle pouvant détecter les distorsions visibles par l'œil humain. L'apprentissage de ces SVM nécessite alors des bases de données munies d'images et des moyennes d'opinions des scores (MOS) correspondantes fournies par les humains.

Les artefacts qui sont concernées vont représenter chacune des classes d'appartenance. Une phase de configuration d'un seuil de détection d'artefacts est nécessaire une fois la SVM entrainée.

A.2.3 Parallélisation et granularité

La parallélisation des SVM effectuées sur de macroblocs d'une image permet une accélération de toute opération de SVM sur cette image. En effet, la faible taille des sous-images constituant ces macroblocs réduit de façon considérable la complexité des algorithmes en termes de temps de calcul. Pour appuyer cette idée, une séquence vidéo nommée "*Fantastic Four theatrical trailer*" en H.264 (par 20th Fox Century) a été partitionnée en trames de 1280 x 544 pixels. Chaque trame a est ensuite été divisée 25 sous-images de taille égale. La valeur moyenne sur les 25 scores relatifs aux 25 sous-images a été comparée par le score de l'image globale afin de vérifier la pertinence de cette idée (voir Figure A.7).

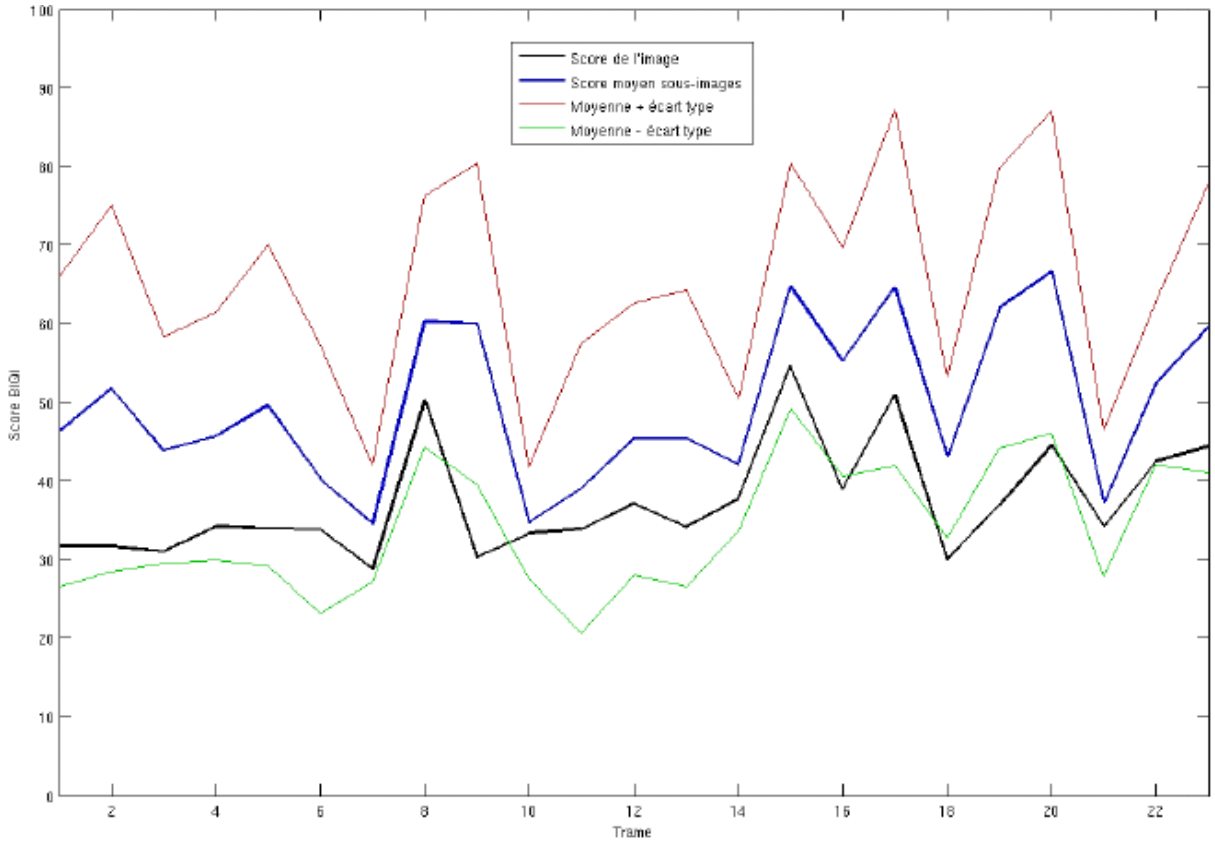


Figure A.0.7 Comparaison entre SVM avec parallélisme et SVM sans parallélisme sur la séq. *Fantastic.*

On peut constater avec la répartition des scores que la moyenne des scores des sous-images reste comprise entre la moyenne des scores des sous-images + ou – l'écart type. Certains scores de sous-images s'écartent du score de la trame ; C'est le cas de la trame 9 qui a un score égal à 30, soit près de la moitié de la moyenne des scores des sous-images.

A.3 Contribution essentielle

La séquence d'image utilisée dans cette expérience est celle de la séquence vidéo *FantasticFour*. Pour diminuer la complexité de l'algorithme, seules quelques trames suffisamment séparées temporellement les unes des autres seront sélectionnées, vu la similarité entre trames voisines. L'expérience vise à développer un détecteur des artefacts flou et effets de bloc JPEG basé sur les SVM. Dans un premier temps seul les effets de bloc sont étudiés.

A.3.1 Détection des blocs JPEG

Vu la nature binaire de la classification, deux classes d'appartenance sont définies: la classe "JPEG" et la classe "non-JPEG". L'utilisation du spectre de puissance de l'image telle qu'illustré en [39] a permis l'élaboration d'un vecteur représentatif adéquat.

a) Vecteur représentatif

L'augmentation du taux de compression sur l'image a pour effet de faire apparaître de façon plus visible les harmoniques de fréquence spatiale des blocs, les autres fréquences étant atténuées par la quantification. On part de cette constatation pour construire un vecteur représentatif (voir Figure A.8) en multipliant les transformées de Fourier discrètes à la fois dans le sens horizontal et vertical du gradient de l'image de façon à renforcer l'apparition des contours des blocs. Puis on effectue un préfiltrage du profil obtenu par la somme des coefficients horizontalement et verticalement. Au final une transformée de fourrier permet de rehausser les limites des blocs.

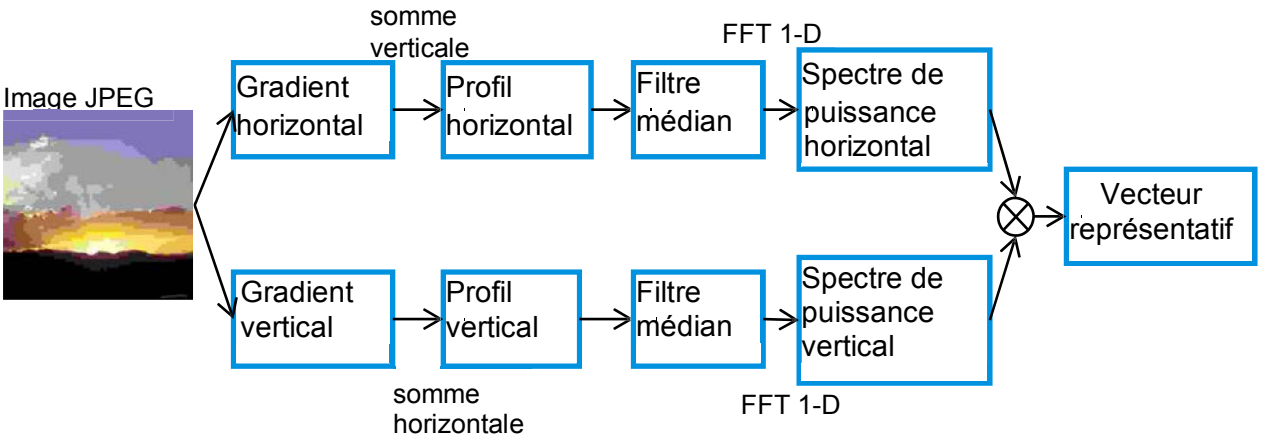


Figure A.0.8 Algorithme de calcul du vecteur représentatif.

b) Choix des jeux de données

Les bases de données utilisées dans cet expérience étaient celles de CSIQ et LIVE.

- CSIQ : 30 images avec 5 niveaux (1 : niveau faible, 5 : niveau élevé) de distorsions à savoir la compression JPEG, la compression JPEG 2000, la baisse globale du contraste, le flou gaussien, ainsi que les bruits gaussiens rose et blanc.

- LIVE : 29 images ayant subi différents niveaux de distorsions à savoir la compression JPEG, la compression JPEG 2000, le bruit blanc, le flou gaussien et la décoloration rapide.

Un équilibre entre les classes non-JPEG et JPEG a été constituées pour homogénéiser le jeu de données de la phase apprentissage. On choisi aussi les échantillons de façon à couvrir différentes valeurs possibles du VS (vecteur support). Les images ayant subi un bruit gaussien rose de la base CSIQ sont ensuite choisies pour l'entraînement. On peut remarquer (voir Figure A.9) que pour cette distorsion, les valeurs du VS balayent tout le spectre.

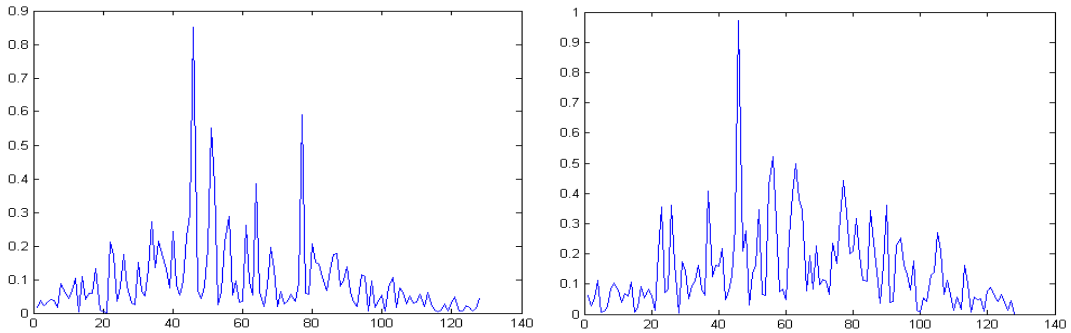


Figure A.0.9 Valeurs du VS (sur l'axe y) pour 2 niveaux (niv. 1 à gche et niv. 5 à dte) du bruit rose sur la même image.

c) **Seuil de classement**

Une fois les vecteurs générés après l'apprentissage et le test, on les associe des classes d'appartenance. Une étude visuelle va permettre de définir un seuil de visibilité des blocs afin de clairement les distinguées des autres effets de la compression JPEG. De ce seuil va dépendre la sensibilité de la SVM. Les images dont le score est inférieur à ce seuil sont classées comme "non-JPEG" et celles dont le score est supérieur classées comme "JPEG".

d) **Outils utilisés**

Une implémentation de la SVM a été faite sur *Matlab*, suivi d'une programmation en C. Un VS représentant une image est généré et enregistré par ligne pour chacun des artefacts JPEG de CSIQ et de LIVE, ainsi que les bruits blanc et rose de CSIQ.

e) **Résultats**

Trois versions ont été réalisées avec différentes tailles du vecteur représentatif.

SVM avec un VS de 24 caractéristiques :

Dans un premier temps la SVM utilise uniquement 24 caractéristiques issues d'une transformée de Fourier sur 24 points, afin d'obtenir une classification rapide pour un contrôle de la vidéo image par image. L'efficacité atteint avec cette première SVM est de 97% en phase apprentissage, et 51.2% en phase de test.

SVM avec un VS de 512 caractéristiques :

Ces résultats peu satisfaisants on conduit à l'élaboration d'une deuxième SVM avec 512 caractéristiques, dimension plus proche de celles de l'image. Avec cette deuxième SVM, efficacité de 77.7% a été atteinte en phase validation et 88.7% en général. Ces performances ont été améliorées après un réglage des paramètres et une optimisation plus précise ainsi qu'une généralisation utilisant une validation avec 20 sous-ensembles. On atteint ainsi une efficacité de 92.4% pour la généralisation et 78.6% en phase apprentissage.

SVM avec un VS de 128 caractéristiques :

Etant donné la grande complexité matérielle d'une FFT sur 512 points, une FFT plus petite a été optée pour les tests. La taille du vecteur a alors été réduite à 128 caractéristiques. On retrouve des performances compétitives, soit 91.1% d'efficacité en validation 96.3% en apprentissage.

A.3.2 Détection de flou

Une métrique de mesure est utilisée comme une composante du VS afin de de consolider la fiabilité de la prédiction. Il importe ainsi d'élaborer la métrique adéquate, ici on opte pour la "métrique objective de flou" (*Objective Blur Metric*) et la métrique CPBD. On s'intéresse notamment à leur monotonie et leur corrélation avec le DMOS (MOS numérisé).

La netteté étant l'aspect contraire du flou dans une image, elle apparait comme une mesure du flou. Une étude a donc été portée sur la mesure de la netteté.

a) **Netteté**

Dans le domaine fréquentiel la netteté est traduite par une faible activité dans les hautes fréquences. En divisant le spectre de puissance de l'image filtrée par un filtre passe-haut (PH) par celui de l'image filtrée par un filtre passe-bande (PB), on obtient un scalaire pouvant constituer une métrique intéressante (voir Figure A.10).

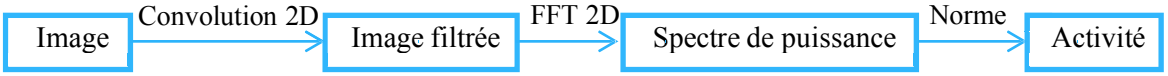


Figure A.0.10 Méthode de calcul de l'activité pour les filtrages Passe-haut et Passe-bande.

Les deux filtres :

$$PH = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{et} \quad PB = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

La netteté des images de la catégorie Flou Gaussien de CSIQ a été mesurée et tracé sur un même graphique (Figure A.11) que leur DMOS. On remarque un une forte corrélation entre la métrique netteté et le DMOS.

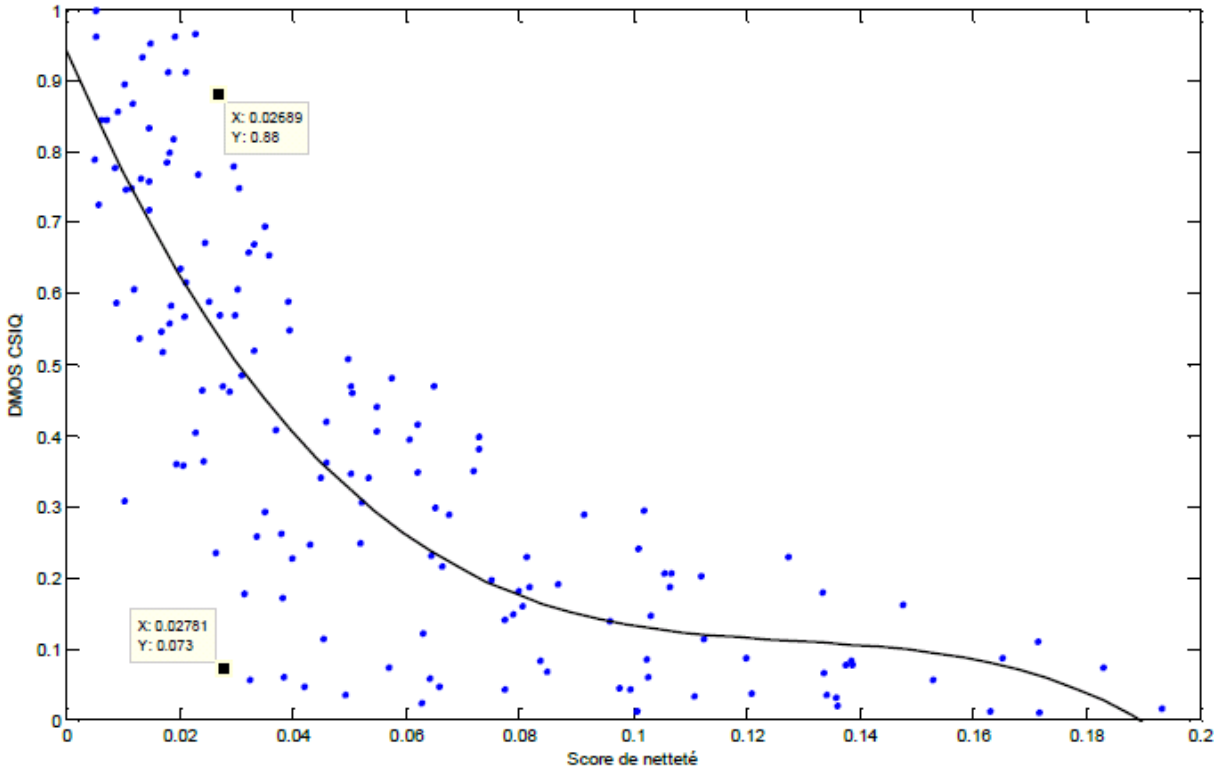


Figure A.0.11 Corrélation entre qualité perceptuelle (DMOS) et la métrique de la netteté.

On obtient un coefficient de corrélation de Spearman de -0.82, qui illustre bien la forte décroissance entre le flou de l'image et la qualité perceptuelle (représentée par le DMOS). Toute fois cette métrique est moins avantageuse à cause de la grande dispersion des points (grande variance).

b) **Métrique objective de flou**

L'élaboration de cette métrique [15] se base sur l'augmentation de dégradation de la qualité visuelle d'une image (voir Figure A.0.12) par l'ajout de flou. On l'obtient donc en faisant la différence entre le degré de flou initial et le degré de flou final d'une image floutée.

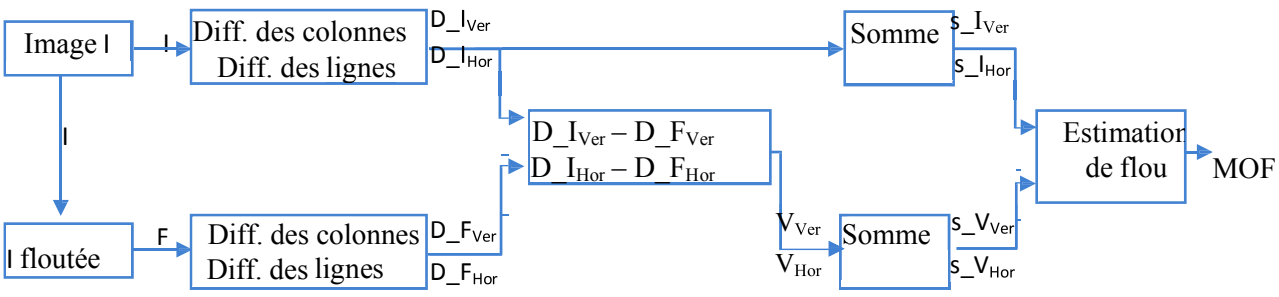


Figure A.0.12 Schéma de calcul de la métrique objective du flou dans une image *I*.

Dans chacun des sens (horizontaux et verticaux) de la matrice image *I*, on réalise un filtrage passe-bas, puis on normalise la somme des variations des coefficients initiaux et finaux pour obtenir un scalaire MOF (Métrique objective du Flou) dans l'intervalle [0 , 1]. La Figure A.13 montre une nette monotonie et une assez forte corrélation entre la métrique du flou et le DMOS qui permettent de considérer cette métrique comme intéressante.

Une étude approfondie a toutefois montré que la métrique ne fonctionne pas pour les images spatialement homogènes, cela se traduit par les points rouges qui s'écartent de l'allure linéaire de la corrélation avec le DMOS.

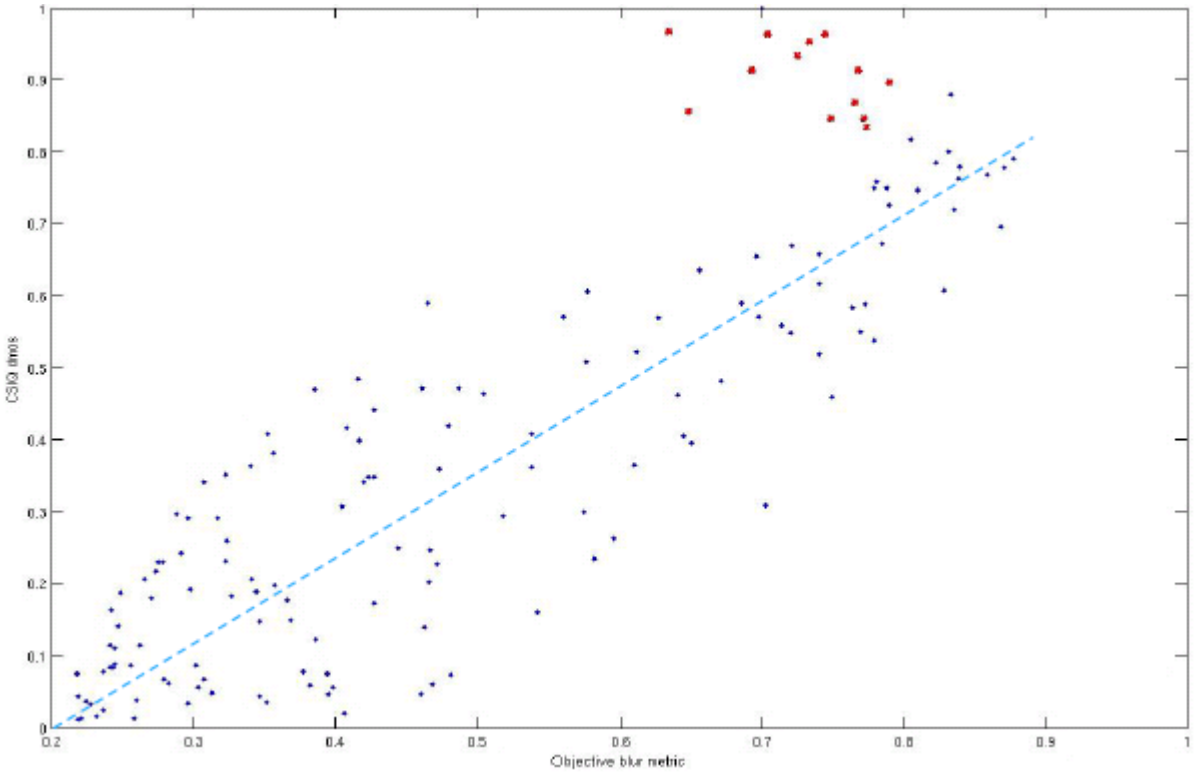


Figure A.0.13 Corrélation entre DMOS et la métrique du flou, base de données CSIQ.

c) **Métrique CPBD**

La métrique CPBD (*Cumulative Probability of Blur Detection*) se base sur le fait que dans une image, le flou se manifeste par des contours moins nets et élabore une méthode appropriée pour leur détection. Cette méthode stipule que le niveau de contraste autour d'un contour définit le niveau la visibilité de ce contour. Un seuil de visibilité de contour sur l'image est défini pour correspondre à un minimum de contraste à partir du quel l'image est dite floue.

La métrique CPBD calcule la probabilité de flou de la manière suivante:

$$P_{flou} = 1 - e^{\left|\frac{largeur\ du\ contour}{seuil\ de\ perceptibilité}\right|^{\beta}} \quad \text{(A.1)}$$

Où β est la variable de normalisation. Elle dépend du contraste global de l'image.

Le seuil de perceptibilité du flou correspond à une probabilité de flou de 63%. En dessous de cette valeur, le flou est indétectable.

La Figure A.14 présente la corrélation obtenue avec le DMOS. On constate une saturation rapide des probabilités de flou à 100%, traduisant une très grande sensibilité de la métrique CPBD. Il est donc nécessaire d'ajuster cette sensibilité par une normalisation (voir Figure A.15). La variance en a été affectée rendant la corrélation avec le DMOS moins évidente à établir.

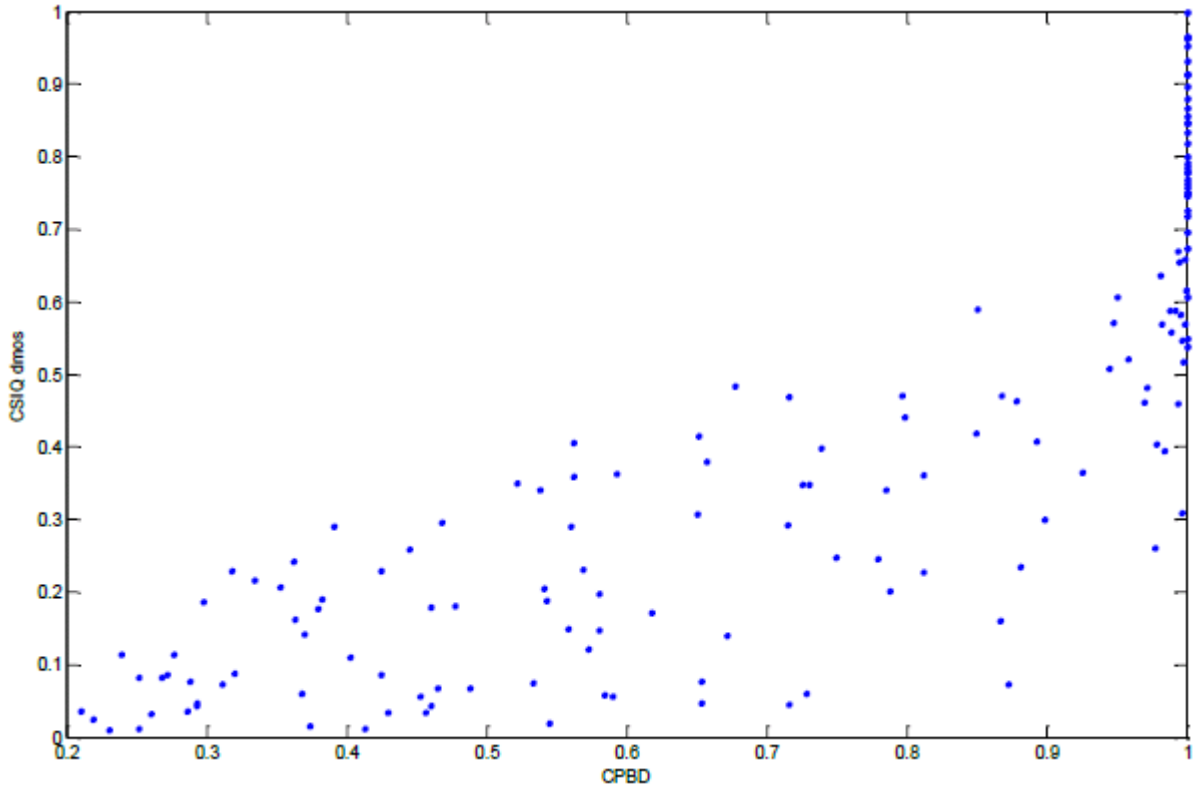


Figure A.0.14 Corrélation entre métrique CPBD et DMOS.

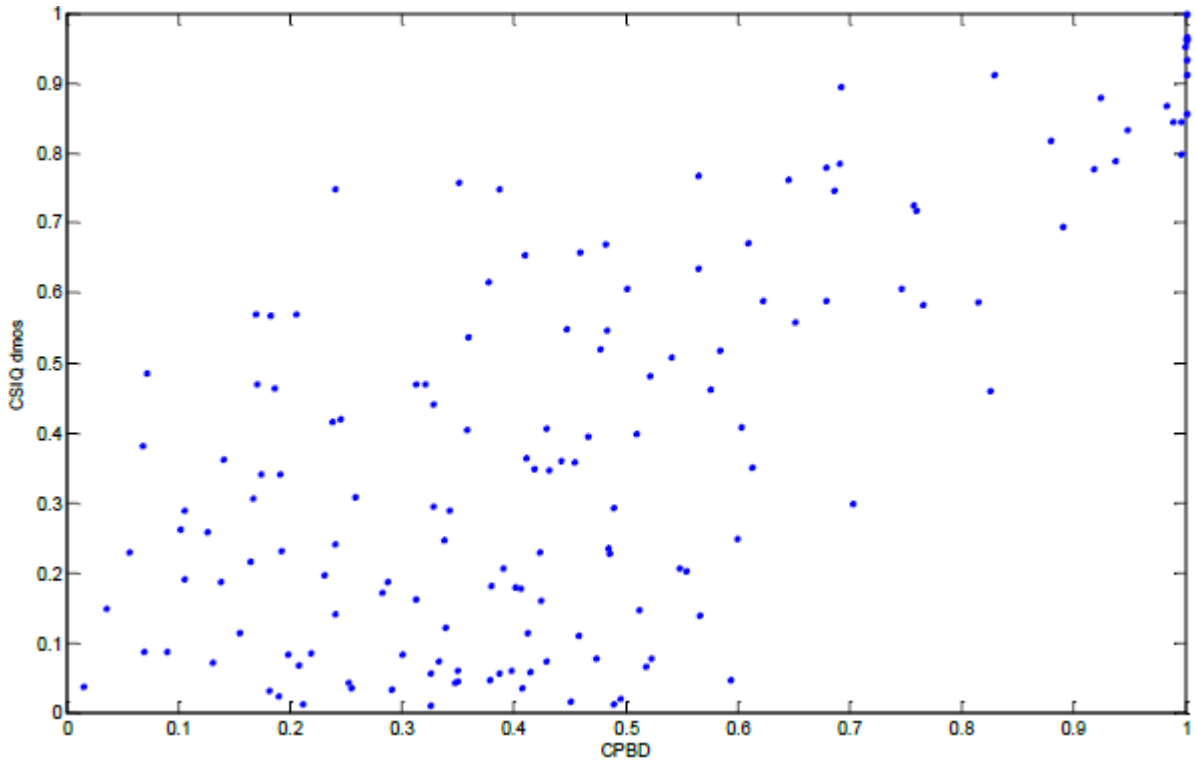


Figure A.0.15 Corrélation CPBD vs DMOS, après normalisation.

A.4 Conclusion

Ce stage réalisé par Mourad GOUMRHAR et en lien avec la présente thèse a permis l'élaboration d'une méthode basée sur les Machines à Vecteurs Support pour la détection du flou et des effets de bloc.

On a montré que cette détection peut s'étendre à plusieurs artéfacts grâce au principe de fonctionnement même des SVM. Il suffit d'intégrer des métriques relatives aux artefacts en question pour former de nouveaux VS (vecteurs représentatifs) représentants toutes les classes associées aux différents artefacts, puis de générer un modèle multi-classes.

De grandes performances ont été atteintes dans la détection des blocs JPEG, grâce à l'élaboration d'un VS basé sur le spectre de puissance. Une efficacité de 92% obtenue lors des tests a montré une bonne capacité de généralisation de la machine. Un challenge a été réalisé quant à l'implémentation bas niveau avec la réduction considérable de la complexité matérielle et des coûts en nombre de transistors et surface d'intégration sur le silicium.

Face à la difficulté de trouver une métrique objective de flou plus générale (en termes de formats d'images, de contenu et de texture) fortement corrélé au DMOS, la combinaison de plusieurs métriques semble être la meilleure alternative.

ISBN : 978-2-11-129194-2

TITLE

ERROR DETECTION AND CONCEALMENT INTEGRATED IN A VIDEO DECODER USING TECHNICS OF STATISTICAL ANALYSIS

ABSTRACT

This report presents the research conducted during my PhD, which aims to develop an efficient algorithm for correcting errors in a digital image decoding process and ensure a high level of visual quality of decoded images. Statistical analysis techniques are studied to detect and conceal the artefacts. A control loop is implemented for the monitoring of image visual quality. The manuscript consists in six chapters. The first chapter presents the principal state of art image quality assessment methods and introduces our proposal. This proposal consists in a video quality measurement tool (VQMT) using the Human Visual System (HVS) to indicate the visual quality of a video (or an image). Three statistical learning models of VQMT are designed. They are based on classification, artificial neural networks and non-linear regression and are developed in the second, third and fourth chapter respectively. The fifth chapter presents the principal state of art image error concealment technics. The latter chapter uses the results of the four former chapters to design an algorithm for error concealment in images. The demonstration considers blur and noise artefacts and is based on the Wiener filter optimized on the criterion of local linear minimum mean square error. The results are presented and discussed to show how the VQMT improves the performances of the implemented algorithm for error concealment.

Keywords: *Statistical analysis, visual quality, digital decoder, Image, Artificial Intelligence, Error concealment, blur artifact, artifact noise.*

TITRE

DÉTECTION ET CONCILIATION D'ERREURS INTÉGRÉES DANS UN DÉCODEUR VIDÉO: UTILISATION DES TECHNIQUES D'ANALYSE STATISTIQUE

RESUME

Ce manuscrit présente les travaux réalisés au cours de ma thèse, dont le but est de développer des algorithmes de correction d'erreurs dans un décodage numérique d'images et d'assurer un haut niveau de la qualité visuelle des images décodées. Des techniques d'analyse statistique sont utilisées pour détecter et dissimuler les artefacts. Une boucle de contrôle de la qualité est implémentée afin de surveiller et de corriger la qualité visuelle de l'image. Le manuscrit comprend six chapitres. Le premier chapitre présente les principales méthodes d'évaluation de la qualité des images trouvées dans l'état de l'art et introduit notre proposition. Cette proposition est en fait un outil de mesure de la qualité des vidéos (OMQV) qui utilise le système visuel humain (SVH) pour indiquer la qualité visuelle d'une vidéo (ou d'une image). Trois modèles d'OMQV sont conçus. Ils sont basés sur la classification, les réseaux de neurones artificiels et la régression non linéaire, et sont développés dans le deuxième, troisième et quatrième chapitre respectivement. Le cinquième chapitre présente quelques techniques de dissimulation d'artefacts présents dans l'état de l'art. Le sixième et dernier chapitre utilise les résultats des quatre premiers chapitres pour mettre au point un algorithme de correction d'erreurs dans les images. La démonstration considère uniquement les artefacts flou et bruit et s'appuie sur le filtre de Wiener, optimisé sur le critère du minimum linéaire local de l'erreur quadratique moyenne. Les résultats sont présentés et discutés afin de montrer comment l'OMQV améliore les performances de l'algorithme mis en œuvre pour la dissimulation des artefacts.

Mots clés: *Analyse statistique, Qualité visuelle, Décodeur numérique, Image, Intelligence artificielle, Correction d'erreurs, Artefact de flou, Artefact de bruit.*